Politics and Governance
Open Access Journal | ISSN: 2183-2463
Volume 8, Issue 2 (2020)

Quantifying Higher Education: Governing Universities and Academics by Numbers

Editors

Maarten Hillebrandt and Michael Huber

COGITATIO

Politics and Governance, 2020, Volume 8, Issue 2
Quantifying Higher Education: Governing Universities and Academics by Numbers

Published by Cogitatio Press
Rua Fialho de Almeida 14, 2º Esq.,
1070-129 Lisbon
Portugal

Academic Editors
Maarten Hillebrandt (University of Helsinki, Finland)
Michael Huber (University of Bielefeld, Germany)

Available online at: www.cogitatiopress.com/politicsandgovernance

Table of Contents

Politics and Governance (ISSN: 2183-2463)

Politics and Governance is an innovative new offering to the world of online publishing in the Political Sciences. An internationally peer-reviewed open access journal, Politics and Governance publishes significant, cutting-edge and multidisciplinary research drawn from all areas of Political Science.

www.cogitatiopress.com/politicsandgovernance

COGITATIO

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2020, Volume 8, Issue 2, Pages 1–5
DOI: 10.17645/pag.v8i2.2585

Editorial

Editorial: Quantifying Higher Education: Governing Universities and Academics by Numbers

Maarten Hillebrandt [1,*] and Michael Huber [2]

[1] Faculty of Law, University of Helsinki, 00101 Helsinki, Finland; E-Mail: maarten.hillebrandt@helsinki.fi
[2] Faculty of Sociology, Bielefeld University, 33615 Bielefeld, Germany; E-Mail: michael.huber@uni-bielefeld.de

* Corresponding author

Submitted: 29 October 2019 | Published: 9 April 2020

Abstract
Over the past decades, 'governing by numbers' has taken a flight in the higher education sector. Performance-based budgeting and quality assurance schemes orient universities to new objectives, while rankings have globalised the metrified observation of higher education at large. Where previously no indicators existed, they are being introduced; where indicators already existed, they are being standardised for purposes of comparison. This thematic issue aims to work towards a more comprehensive understanding of the growing diversity of quantification-based instruments in higher education sectors in three European countries. The effects of quantification are noticed at all levels of the higher education system, from policy makers at the top of the regulatory pyramid down to students and academic staff. Yet even quantifiers outside of the regulatory system, such as ranking and metrics organisations, may have an important bearing on the operation of the university organisation and the sector at large. Thus, an entire governance landscape emerges in which actors at various levels turn to numbers for guidance. The articles in this thematic issue analyse the life cycle of such numbers, from their origins, through to their production and finally, their consequences. This editorial outlines the central questions and overarching issues addressed by the thematic issue and introduces its various contributions.

Keywords
comparative policy studies; higher education governance; managerialism; performance indicators; quantification; regulation

Issue
This editorial is part of the issue "Quantifying Higher Education: Governing Universities and Academics by Numbers" edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

1. Public Sector Quantification in the Limelight

Twenty-five years after Porter (1995) focussed attention on the centrality of numbers in the conduct of modern social and political life, the place of quantification in the limelight of the social sciences appears undiminished. Indeed, as overview articles by Espeland and Stevens (2009) or Popp Berman and Hirschman (2018) have shown, over time, quantification research has increasingly branched out, become institutionalised, and settled on a division of labour.

One of the critical assumptions of quantification studies or 'governance by numbers' is the idea that numbers transform organisational and political behaviour. Two powerful time diagnoses by Power (1997) and, more recently, Dahler-Larsen (2012) have argued that late modern society is characterised by relentless efforts at audit and evaluation. This trend, in a memorable phrase by Miller (2001, pp. 381–282), has generated an "avalanche of numbers" for decision making in "almost any organization." This has had remarkable effects. As early as 1999, Hood, Scott, James, Jones and Travers, in a landmark study, estimated that policies for control inside government cost £1 billion per year in the UK alone (Hood et al., 1999, p. 42). Since then, the scale of quality-controlling, competition-inducing, and waste-watching in the pub-

lic sector seems only to have grown, although accents may have changed and intensities differ across countries (Hood, 2007; Hood, James, Peters, & Scott, 2004). Thus, the promise of governance by numbers not only comes with a price tag, but also alters routines and certainties; and that requires academic scrutiny.

It is in this context that this thematic issue turns its attention to the advent of quantification in the higher education sector where 'governing by numbers' appears to have taken a flight. Performance-based budgeting and quality assurance schemes orients universities to new objectives (Ferlie, Musselin, & Andresani, 2008; Huber & Hillebrandt, 2019), while rankings have globalised the metrified observation of higher education at large (Espeland & Sauder, 2007, 2016). Where previously no indicators existed, they are being introduced; where indicators already existed, they are being standardised for purposes of comparison. The proliferation and diversification of quantification has already been diagnosed; the next step is to analyse its manifestations and effects in the fields it is applied to.

2. Puzzles Addressed in This Thematic Issue

This thematic issue is the result of a workshop on quantification in higher education that was held at Bielefeld University in March 2019. Together, the articles present empirical research from three western European countries (England, Germany, the Netherlands), covering various manifestations of quantification, including rankings, finance models, quality assurance, and performance analytics. In spite of their thematic and conceptual diversity, the various contributions are united by a number of common concerns and recurrent themes at the core of quantification studies. We would here like to highlight two of them.

First, the contributions provide further insight into what could be labelled 'the life cycle of quantification', i.e., covering instruments from their origins, through to their production, and finally, their consequences. Starting off from a perceived weaving fault in the existing system, policy makers typically develop rudimentary systems of indicators and targets in order to clarify lines of accountability, increase efficiency, or enhance the innovative potential of universities. Problems often emerge when quantification instruments run into the complexities of a differentiated regulatory field. In higher education, this encompasses the multiple tasks (teaching/research) and foci (fundamental/applied research, different disciplines) of universities, as well as the differential systems within which they and their staff operate (tenured/non-tenured positions, national educational systems). This typically casts different groups within the system against each other. Depending on the situation at hand, this may be universities against ministries, professors against junior academics, or resource-intensive against student-intensive disciplines. The consequences of these quantification-induced conflicts depend on the question how much 'damage' a system is willing to accept. Here, there appear to be marked differences depending on national legacies, with the German and Dutch system more careful and the English system more far-reaching.

A second theme of the articles concerns the 'regulatory nexus in higher education.' Regulators at ministries and university administrations, confronted with political and societal pressures as well as international norms and standards, are thought to fall back on quantification in order to raise their profile in the competition for scarce resources, be it money, students, or prestige. In this account, faculties and academic staff are the largely passive but resisting recipients of their administration's and external regulators' steering strategies (Schimank, 2005; Strathern, 2000). The contributions in this thematic issue question the schematic nature of this account, sketching a more dynamic picture of interactions between policy makers and professionals. The former are often recruited from the latter and show a reluctance to implement too far-reaching quantification instruments, while professionals themselves often complain about inefficiency and a lack of tangible recognition of their achievements. Moreover, several contributions raise attention to the pivotal role of newly emerging economic actors outside of the regulatory system, such as ranking and metrics organisations. The new calculative infrastructures that these actors create are critical in the understanding both of the diffusion of numbers as well as specific constraints to governance.

3. The Articles in This Thematic Issue

The contributions to the thematic issue are clustered around these two themes. The articles by Kandiko Howson and Buckley (2020), Dix, Kaltenbrunner, Tijdink, Valkenburg, and de Rijcke (2020), and Huber (2020) take up aspects of the life cycle theme and offer case studies of quantification in each of the three countries illustrating how governance by numbers is designed, and how the various actors respond to signals of dysfunctionality that emerge over time. The regulatory nexus theme includes articles by Ringel, Brankovic, and Werron (2020), Hillebrandt (2020), and Krüger (2020). These authors address the emergence and usage of new numerical data infrastructures, paying special attention to the administrative and organisational capacities that these infrastructures enable, but also require.

Kandiko Howson and Buckley (2020) describe the career of quantification as a way to measure learning gain in UK undergraduate education. From the 1990s onward, attempts by government regulators to create a market for higher education required an informational level playing field, which quantification was meant to deliver. Since tuition fees nearly trebled to £9,000 per annum in 2012, the drive to give students demonstrable 'value for money' increased. Such value, the authors point out, is routinely defined in terms of "'corporate culture' [for

universities] and individual monetary gain [for students]" (Kandiko Howson & Buckley, 2020, p. 7). In this context, measuring learning quality logically makes up the latest incarnation of this marketisation agenda. The authors offer a detailed account of the vast efforts and material resources that the UK regulator has expended on translating this ideal into reality. Several parties were included in pilot projects for the development of suitable indicators. The efforts began to be resisted when affected actors noticed that indicators were borrowed from the rather dissimilar US context, and were going to be applied with little heed to disciplinary and other variations within universities. As specific policies faltered, the data remained in place. In the end, a situation emerges of numbers in search of a purpose. Or, in the authors' words, "without a rationale for developing, selecting and using measures the number…becomes an end in itself" (Kandiko Howson & Buckley, 2020, p. 11).

The contribution of Dix et al. (2020) focusses on the Dutch context. Their case consists of research departments within a medical centre which are confronted with a new performance measurement instrument that directly affects their funding situation. Like Kandiko Howson and Buckley (2020), the case emerges against a backdrop of marketisation. However, the author's approach to the concept of 'value' is rather different. Borrowing from the notion of 'economies of worth' coined by Boltanski and Thévenot, the authors analyse a controversy over quantification with quite real consequences: The question whether a 'market language' of value should determine the allocation of scarce resources. The analysis lays bare the 'epistemic language' by which opponents of market-based competition criticise the quantification instrument in place. The strength of this article lies in its detailed analysis of sixteen interviews conducted with actors across the organisation, which allows for a nuanced account. While quantification is proposed as a way of trading arbitrary finance inequalities for a more transparent system, management also proves to be sensitive to some of the discontents voiced by opponents of marketisation. For example, it is unwilling to put the long-term financial viability of departments at risk. The article thus shows that arguments for and against quantification-based marketisation can get blended. In the authors' words, "we should be careful in depicting higher education as populated by 'market universities' fully enmeshed in 'epistemic capitalism"' (Dix et al., 2020, pp. 23–24).

The article by Huber (2020) complements the previous two contributions by addressing the role of financial quantification in a German university. University management is represented as a team charged with bridging the expectations gap between state regulators and academics. The role of quantification in this balancing act is inherently fragmented and uneven: Rather than steering the university systematically towards a unified goal, various types of numbers introduced at different moments in time address a multiplicity of audiences within the university organisations. This is not accidental. Rather, multiple policies are in place to remedy the perverse effects of other policies, leading to a conceptual representation of university quantification as foremost an organic, layered whole. Here, the author refers to the organisational sociology of Luhmann: "reforms do not form part of a grand plan, but recurrently intervene to repair the shortcomings and unintended effects of previous interventions" (Huber, 2020, p. 26). The dynamic and multicentred model of quantification allows for a number of interesting conclusions. It appears clear that reality of quantification in practice is inherently a 'messy affair' without one single premeditated logic. Yet, at the end of the day, the layering of quantification succeeds, however transiently, in creating a certain stability of purpose: management by numbers "can to a large extent be incorporated into the traditional organisational form of the university" (Huber, 2020, p. 34).

The contribution by Ringel et al. (2020) shifts the focus in several respects. Rather than the effects of quantification, it considers its preconditions. The selected case, that of higher education rankings, constitutes a form of quantification that has over the past decades increasingly 'gone global' (see also Brankovic, Ringel, & Werron, 2018). The authors bring attention to an aspect that has hitherto received scarce attention in this lively research field: The material preconditions for the creation and sustenance of rankings. A key reason for the 'rankings boom' in higher education and other areas, the authors argue, lies in the organisational turn. They demonstrate how organisations offer capacities for the creation and dissemination of ranking that vastly exceed those of individuals. As a result, ranking organisations are able to publish increasingly sophisticated rankings with a regular frequency, while offering the 'maintenance work' required to bind them to ever-more diverse audiences of 'consumers.' Thus, as the authors point out, "the organizational production of rankings provides an elementary and hitherto overlooked infrastructure" accounting for their success and pervasiveness (Ringel et al., 2020, p. 44). The article's focus offers an interesting additional perspective by 'turning the camera' on a group of actors that fall outside of the regulatory context proper, but which nevertheless have gathered considerable clout in parts of the higher education sector, in a trend that has simultaneously emerged in other competitive sectors such as hospitality, catering, and health care.

In his article, Hillebrandt (2020) observes that although gathering quantified data is a growing business everywhere, in Germany actual reliance on this data is limited. What explains the limited effect of regulatory control through numbers in Germany? With the catchy metaphor of the Mercedes that is left in the garage, the author explores three hypotheses: First, a legal hypothesis that suggests quantification being curbed by legal protections of higher education providers: The state wants 'to drive the Mercedes,' but legal protections wreck this strategy. The state gives up on steering through numbers

while it continues to collect data. A second, labelled dysfunctionality hypothesis holds that regulators see quantification as a flawed and impracticable pursuit. The Mercedes may be shiny, but cannot be driven safely. The flaws and inaccuracies of steering by numbers are considered too momentous to risk failure. The third hypothesis reflects the federal structure of Germany and suggests that federal comparison exposes differences and thus potential weaknesses of the responsible *Länder* (German states): as such, the Mercedes better stays in the garage. The article finds that all three hypotheses contribute to an explanation of the German higher education sector's engagement with quantification, but it also shows that the Mercedes is driven sub rosa, meaning that any governing by numbers "functions in a largely tempered, hybrid, and untransparent manner" (Hillebrandt, 2020, p. 55).

The article by Krüger (2020) forms a suitable concluding note to the empirical section of this thematic issue. In her contribution, Krüger (2020) goes in search of 'quantification 2.0' in the amalgam of commercially exploited data infrastructures that have begun to emerge out of increasingly vast bibliometric data sets. Beyond marketisation and managerialism, she argues, students of quantification in the higher education sector should not overlook ongoing technological development as a push factor in its own right. As the author states, "bibliometric data has turned into a self-serving end while their providers are constantly seeking for new tools to make use of them" (Krüger, 2020, p. 59). Organisations in the higher education sector and beyond have begun to collect data without any specific purpose, under the (ideologically informed) supposition that eventually uses will be found for it that will improve organizational performance. New functionalities invite academics, managers, and regulators to engage with—and even create—data in novel ways, thereby fomenting alternative ways of regarding academic work in relation to the individual and (a reinvented notion of) the wider profession. Does all of this spell a dystopian future in which untransparent private firms remake the academic profession in their commercial image? The author remains cautious here, by pointing out that in spite of its performativities, the various new potential uses of 'quantification 2.0' are bounded by notions of customer (academic community) acceptance among developers.

Finally, a commentary by Hamann (2020) reflects on the findings in the articles that form part of this thematic issue, focussing in particular on the sociological implications of higher education governance by numbers. Setting out from the Foucauldian dyad of power and discipline, he places the phenomenon of academic quantification primarily in the light of what he describes as 'panopticism': The creation of new avenues for number-based mutual observation that rearrange interactions and create new forms of control. 'Numerocratic panopticism' however reveals certain dynamics that depart from the classic panoptic gaze: it increases the number of observers, is freed from spatial constraints, and its normative programme is more open-ended. This, the author argues, demonstrates that "governance by numbers is not only an epitome of classical panopticism but…also a panopticon reversed" (Hamann, 2020, p. 70).

All in all, the various contributions offer a glimpse onto a wider higher education governance landscape oriented on numbers for guidance. In doing so, they show a considerable attention to the details and intricacies of specific policy instruments, and the way in which they come about. As it turns out, such attention to detail often pays off as it highlights where problems occur and thus, how the fault lines of policy conflict are likely to play out. It becomes clear from the various empirical studies that in most instances of quantification, such problems do emerge. Typically, quantification-based steering instruments require fundamental reform within a few years after their introduction. We attribute this general observation to the short-term focus of many indicators, which means that their dysfunctionalities take some time to become apparent. Yet, the contributions to this thematic issue inevitably lead to the conclusion that, even when repeatedly bruised and floored, the appeal of higher education quantification continues undiminished.

Acknowledgments

The authors would like to thank the participants to the workshop "Quantifying Higher Education: Origins, Production, Consequences," that was held at Bielefeld University on 28–29 March 2019. The authors further gratefully acknowledge funding received from the Deutsche Forschungsgemeinschaft (DFG) (Project No. 627097) under the Open Research Area Scheme (Project Title: QUAD—Quantification, Administrative Capacity and Democracy, 2016–2019).

Conflict of Interests

The authors declare no conflict of interests.

References

Brankovic, J., Ringel, L., & Werron, T. (2018). How rankings produce competition: The case of global university rankings. *Zeitschrift Für Soziologie*, *47*(4), 270–288.

Dahler-Larsen, P. (2012). *The evaluation society*. Redwood City, CA: Stanford University Press.

Dix, G., Kaltenbrunner, W., Tijdink, J., Valkenburg, G., & de Rijcke, S. (2020). Algorithmic allocation: Untangling rival considerations of fairness in research management. *Politics and Governance*, *8*(2), 15–25.

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, *113*(1), 1–40.

Espeland, W. N., & Sauder, M. (2016). *Engines of anxiety: Academic rankings, reputation, and accountability*. New York, NY: Sage.

Espeland, W. N., & Stevens, M. L. (2009). A sociology of quantification. *European Journal of Sociology*, *49*(3), 401–436.

Ferlie, E., Musselin, C., & Andresani, G. (2008). The steering of higher education systems: A public management perspective. *Higher Education*, *56*, 325–348.

Hamann, J. (2020). Governance by numbers: A panopticon reversed? *Politics and Governance*, *8*(2), 68–71.

Hillebrandt, M. (2020). Keeping one's shiny Mercedes in the garage: Why higher education quantification never really took off in Germany. *Politics and Governance*, *8*(2), 48–57.

Hood, C. (2007). Public service management by numbers: Why does it vary? Where has it come from? What are the gaps and the puzzles? *Public Money and Management*, *27*, 95–102.

Hood, C., James, O., Peters, B. G., & Scott, C. (Eds.). (2004). *Controlling modern government: Variety, commonality and change*. Cheltenham: Edward Elgar.

Hood, C., Scott, C., James, O., Jones, G., & Travers, T. (1999). *Regulation inside government: Waste-watchers, quality police and sleaze-busters*. Oxford: Oxford University Press.

Huber, M. (2020). Steered by numbers: How quantification differentiates the reform of a German university. *Politics and Governance*, *8*(2), 26–35.

Huber, M., & Hillebrandt, M. (2019). "Pay for promise" in higher education: The influence of NPM on resource allocation in German universities. *Historical Social Research*, *44*(2), 247–269.

Kandiko Howson, C., & Buckley, A. (2020). Quantifying learning: Measuring student outcomes in higher education in England. *Politics and Governance*, *8*(2), 6–14.

Krüger, A. K. (2020). Quantification 2.0? Bibliometric infrastructures in academic evaluation. *Politics and Governance*, *8*(2), 58–67.

Miller, P. (2001). Governing by numbers: Why calculative practices matter. *Social Research*, *68*(2), 379–396.

Popp Berman, E., & Hirschman, D. (2018). The sociology of quantification: Where are we now? *Comparative Sociology*, *47*(3), 257–266.

Porter, T. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Power, M. (1997). *The audit society*. Oxford: Oxford University Press.

Ringel, L., Brankovic, J., & Werron, T. (2020). The organizational engine of rankings: Connecting "new" and "old" institutionalism. *Politics and Governance*, *8*(2), 36–47.

Schimank, U. (2005). New public management and the academic profession: Reflections on the German situation. *Minerva*, *43*, 361–376.

Strathern, M. (Ed.). (2000). *Audit cultures: Anthropological studies in accountability, ethics and the academy*. London: Routledge.

About the Authors

Maarten Hillebrandt is Postdoctoral Researcher at the Erik Castrén Institute, University of Helsinki, where he is involved in a research project on the transparency of EU decision making. Previously, he participated in the international comparative ORA-funded research project 'Quantification, Administrative Capacity and Democracy' (2016–2019), in which researchers from the London School of Economics, Paris MinesTech, Bielefeld University, Helmut Schmidt University Hamburg, and Leiden University were co-collaborators. Maarten specialises in EU institutional law and politics, government transparency policy, and quantification.

Michael Huber is Professor of Sociology at the Department of Sociology at Bielefeld University. His research is focused on organisational sociology, higher education studies, and regulatory policymaking, more specifically, risk-based regulation.

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2020, Volume 8, Issue 2, Pages 6–14
DOI: 10.17645/pag.v8i2.2564

Article

Quantifying Learning: Measuring Student Outcomes in Higher Education in England

Camille Kandiko Howson [1,*] and Alex Buckley [2]

[1] Centre for Higher Education Research and Scholarship, Imperial College London, London, SW7 2AZ, UK; E-Mail: c.howson@imperial.ac.uk
[2] Learning and Teaching Academy, Heriot-Watt University, Edinburgh, EH14 4AS, UK; E-Mail: alex.buckley@hw.ac.uk

* Corresponding author

Submitted: 17 October 2019 | Accepted: 25 November 2019 | Published: 9 April 2020

Abstract
Since 2014, the government in England has undertaken a programme of work to explore the measurement of learning gain in undergraduate education. This is part of a wider neoliberal agenda to create a market in higher education, with student outcomes featuring as a key construct of value for money. The Higher Education Funding Council for England (subsequently dismantled) invested £4 million in funding 13 pilot projects to develop and test instruments and methods for measuring learning gain, with approaches largely borrowed from the US. Whilst measures with validity in specific disciplinary or institutional contexts were developed, a robust single instrument or measure has failed to emerge. The attempt to quantify learning represented by this initiative should spark debate about the rationale for quantification—whether it is for accountability, measuring performance, assuring quality or for the enhancement of teaching, learning and the student experience. It also raises profound questions about who defines the purpose of higher education; and whether it is those inside or outside of the academy who have the authority to decide the key learning outcomes of higher education. This article argues that in focusing on the largely technical aspects of the quantification of learning, government-funded attempts in England to measure learning gain have overlooked fundamental questions about the aims and values of higher education. Moreover, this search for a measure of learning gain represents the attempt to use quantification to legitimize the authority to define quality and appropriate outcomes in higher education.

Keywords
accountability; education; governance; learning; quality assurance

Issue
This article is part of the issue "Quantifying Higher Education: Governing Universities and Academics by Numbers" edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

1. Introduction

Since 2014, the government in England has undertaken a programme of work to explore the measurement of learning gain in undergraduate higher education, defined for the purposes of the programme as "a change in knowledge, skills, work-readiness and personal development, as well as enhancement of specific practices and outcomes in defined disciplinary and institutional contexts" (Kandiko Howson, 2019, p. 5). This is part of a wider neoliberal agenda in England, as over the past decade the government has driven the development of a competitive market in higher education (Naidoo & Williams, 2015; Olssen, 2016). Browne (2010) suggested new forms of financing higher education and supporting widening participation (2010), with the Department for Business, Innovation and Skills moving to put 'students at the heart of the system' through shifting the burden of funding more completely from grants to tuition fees, and from the state to students (2011); home student fees trebled to £9,000 per year in 2012 under the leadership of the Minister of State for Universities and

Science David Willetts. A competitive market was fully put in place through the removal of student number allocation and the complete uncapping of student numbers by the Treasury in 2015. A market for students—with associated neoliberal ideology of a subsequent increase in quality—was designed, linking teaching excellence, social mobility and student choice (Department for Business, Innovation and Skills, 2016). This was implemented through new managerialism within higher education with a focus on outputs such as rankings to drive competition within a neoliberal market (Lynch, 2015).

Under neoliberal logic, to support a competitive market there is a need for information on how institutions are performing. Given the thousands of courses across hundreds of diverse institutions there is intense subjectivity in how 'excellence' is understood; however, quantification of performance gives the "appearance of scientific objectivity" (Ehrenberg, 2003, p. 147). This provides rankings and frameworks with their credibility as resources of information and as arbiters of value for higher education.

This neoliberal agenda understands 'value' primarily in terms of "corporate culture" and individual monetary gain (Giroux, 2002, p. 429), with student outcomes featuring as a key construct of value for money for students, alongside value for money for the state. These notions of value are increasingly subjected to measurement. However, a perennial question of social science research remains: are those meaningful concepts of value? And if they are not, what is the value of the measure? The assessment of learning gain started as a debate about the benefits that students were accruing from their time and investment in higher education. However, those more fundamental questions about quality have been lost in a search for quantity—the need for a numerical representation of quality, even if divorced from what it represents. In this article we explore the issues raised by the process of quantification represented by the learning gain initiative, particularly around who decides what students should learn, what higher education is for and how its value is measured. We suggest that the recent search for measures of learning gain in the UK is an example of a shift from quantification as a mechanism for representing value, to quantification becoming the value itself.

2. Interest in Large-Scale Learning Metrics

A range of evidence has prompted concerns about the value of what students derive from their investment in higher education, mostly out of the US due to escalating tuition fees and practices of for-profit providers. Research from the US indicates that there is a gap between employers and graduates' views on the level of achievement of essential employability skills (Hart Research Associates, 2015), and varying conceptions of employability skills across stakeholders (Tymon, 2013). There is debate about the role of using employability metrics in higher education outcomes, particularly in relation to generic outcomes as employers often have specific skill requirements from graduates (Cranmer, 2006; Frankham, 2016). A high-profile study in the US using the Collegiate Learning Assessment (CLA) instrument to explore what students are gaining from higher education seemed to find that significant proportions of students are not developing key skills such as critical thinking and complex reasoning (Arum & Roska, 2011). This raised questions about what students were learning and whether it was 'enough.'

This question was at the heart of an Organisation for Economic Co-operation and Development feasibility study, the Assessment of Learning Outcomes in Higher Education. It was run across multiple countries and subjects of study. However, it faced challenges around questions of what to measure, with international, cultural and subject-level differences emerging. Due to concerns about data quality and use, the project was not continued (Organisation for Economic Co-operation and Development, 2013a, 2013b). This project identified the challenge of trying to develop a generic instrument across different disciplinary and national contexts.

The findings from the US and questions being asked globally resonated in the UK, which faced extensive political debates and student protests about raising tuition fees, alongside concerns about 'grade inflation' promoted by rises in the awarding of first-class degrees (Bachan, 2017). As a complement to changing the funding system to promote a market culture in higher education, the Minister David Willetts identified a need for comparable information to promote student choice and for accountability of the large sums of student fees entering the system, backed by public loans.

Existing global rankings such as those produced by the Times Higher Education use quantification as the basis of quality, (Hazelkorn, 2015) but focus on research and reputation. In the UK, the domestic rankings, compiled by major newspapers, include measures of student satisfaction drawn from the National Student Survey. However, the National Student Survey does not attempt to directly measure student learning, and there has been very little effort to establish a correlation between the National Student Survey scores and successful learning; a rare recent study suggests they may in fact be inversely related (Rienties & Toetenel, 2016).

In the 1990s in the US, a similar lack of large-scale data related to student learning was noted, alongside the rising importance of research and reputation-based rankings. This led to development of the National Survey of Student Engagement, which is a distillation of decades of evidence on what activities promote student success (retention, progression and completion) into items which provide actionable data for students and staff (e.g., asking questions in class, such as 'Do students do this?' or 'Can staff provide more opportunity for this to happen?'). It also provides benchmarked data and has a well-developed evidence-base for enhancing teaching and promoting student learning. It is now used across

the world (Coates & McCormick, 2014), and although a version has been developed for use in the UK (Kandiko Howson & Buckley, 2017), it has had relatively limited impact due to competition from the nationally-mandated National Student Survey.

The challenges encountered by international efforts to measure student learning, and associated outcomes such as graduate employability, show the dominance of national issues in higher education policy making. Even when schemes such as the UK's Research Excellence Framework are adopted by other countries the policies are adapted locally and not used comparatively (see the Excellence in Research for Australia, 2018). Efforts to measure student learning are bounded by cultural, structural and institutional differences across countries. Different conceptual definitions and student populations mean many data elements are not comparative (Matsudaira, 2016). For example, international students are variously seen in a deficit model, as taking local places, as a drain on public services or as a financial benefit (see Kandiko Howson & Weyers, 2013). Without international benchmarks in place, however, national efforts to measure student learning are highly politicised, as they are costly to design and administer. To justify the substantial investment, initiatives need to show the value both of the development of measurement tools, and—for political reasons—of national higher education sectors.

3. Origin of Measures of Learning Gain in England

Through the political desire to create a competitive market in higher education (Department for Business, Innovation and Skills, 2011), the actions of various policy actors and global influences, the government in England embarked on a large-scale effort to measure student learning gain. The initial catalyst for the learning gain agenda was the changes to tuition fee structure and the identification by the Minister of a lack of information for students to make 'value' decisions about what and where to study. As an indication of the policy complexity, the work was originally driven by three sector bodies that no longer exist: the Department for Business, Innovation and Skills (whose University remit moved to the Department for Education in 2016) alongside the Higher Education Funding Council for England (whose activities were taken over by the new regulator, the Office for Students in 2018) and the Higher Education Academy (which merged into AdvanceHE in 2018). Work started with a scoping study which developed a definition of learning gain as "the 'distance travelled' or the difference between the skills, competencies, content knowledge and personal development demonstrated by students at two points in time" (McGrath, Guerin, Harte, Frearson, & Manville, 2015, p. xi). This broad, generic view of learning gain contrasted with the academic literature, which defines it more narrowly, for instance as "the academic and personal transferable attributes gained as a result of the active pursuit of content-specific knowledge in a given course of study" (Coates & Mahat, 2014, p. 17).

In 2015, the Higher Education Funding Council for England then led on designing three strands of activity to test various methodological approaches to measuring learning gain. Firstly, there was a suite of 13 pilot projects involving over 70 institutions. A second area focused on analysis of existing government databases to explore the possibility of finding proxy measures of learning gain. The third strand was initially mooted as developing a standardised assessment for students, however after backlash from the sector this was reconceptualised as a project based on the Wabash National Study led by the Center of Inquiry (2016). The Wabash project was a large longitudinal study which used multiple process and output measures to explore the impact of liberal arts education on student learning across multiple institutions in the US (Pascarella & Blaich, 2013).

However, as the strands developed, it was not clear to stakeholders what was being measured, or why, compounded by changes at the Ministerial level which resulted in a lack of intellectual leadership of the agenda. The Higher Education Funding Council for England provided an amended definition of learning gain on its website when the projects were launched, as "an attempt to measure the improvement in knowledge, skills, work-readiness and personal development made by students during their time spent in higher education" (2018, p. 1). Most of the pilot projects developed their own working definition of learning gain, referenced in project webpages (Higher Education Funding Council for England, 2018), such as The Open University-led project adopting "a growth or change in knowledge, skills, and abilities over time that can be linked to the desired learning outcomes or learning goals of the course" the University of Lincoln-led project using "the extent to which undergraduate students have gained a key set of transferable skills and competencies that prepare them for the next stages of their career upon graduation, be it employment or further study" and "the extent to which participating in work-based learning, or work preparation activities, contributes to the readiness of the graduate to participate in a professional context" by the Ravensbourne-led project. These varied definitions indicate the complex territory of learning gain and the lack of consensus over what 'counts' as learning gain; measures are not neutral; they define what matters (Lynch, 2015; Power, 1994).

This led to debate across the sector about what constitutes a learning gain measure, with learning gain becoming an umbrella term for a wide variety of indicators relating to the student experience and student outcomes. There was further confusion with the development of the Teaching Excellence Framework, led by the Department for Business, Innovation and Skills, which aimed to assess teaching excellence and to adopt principles of quality-based funding, with 'Student Outcomes and Learning Gain' as one of the three pillars of quality explored (Gunn, 2018). Although technically separate policy initiatives,

there was extensive speculation in whether the learning gain programme would develop an outcomes metric that could be used for institutional comparison linked to funding. Furthermore, when taking over from the Higher Education Funding Council for England part-way through the learning gain projects, the Office for Students set itself up as a data-driven regulator, but without a clear position on future plans for learning gain.

Due to a lack of leadership of the initiative, the various sector stakeholders could not agree whether a use for the metrics should come first, such as designing institutionally comparative measures to measure performance and provide accountability, or whether valid measures of learning gain needed to be developed, that then could potentially be used for a variety of purposes, including enhancing teaching and learning and assuring quality. The projects struggled to develop measures without a clear direction for what they would be used for, as this impacts how measures are designed.

Whilst valid measures in specific disciplinary or institutional contexts were developed, such as concept inventories in Chemistry and mathematical models for institutions delivering higher education in further education settings (Kandiko Howson, 2019), a robust single instrument or measure failed to emerge. It also became apparent that the metrics devised were not as straightforward as hoped for by policymakers. Even existing measures such as students' grades demonstrated wide discrepancies across modules, courses and institutions.

The programme of work was beset with challenges of student engagement and interrelated issues around data protection, data sharing and research ethics. These challenges stemmed from a lack of rationale or clear purpose for measuring and using the data. Indeed, "The greatest challenge in developing learning indicators is getting consensus on what kind of learning should be measured and for what purpose a learning indicator is to be used" (Shavelson, Zlatkin-Troitschanskaia, & Mariño, 2018, p. 251). The focus on developing measures, rather than what needs measuring and why, has led to a circular policy development model rather the usual uni-directional causal model (Birkland, 2015). The outcomes of the learning gain programme became a solution in search of a problem. They successfully identified disciplinary-level differences both in terms of absolute outcomes but also in terms of what was valued, such as what successful communication skills are in Medicine and Law, and the role of reflection in Humanities and pre-professional subjects. However, government policy and regulatory levers operate instead at the institutional level.

4. Learning Gain and the Disciplines: US and UK Examples

The projects identified the discipline as the primary unit of comparison for student learning outcomes in England. However, policymakers were interested in a generic instrument which could be used to compare institutions, which became the focus of the two other strands of activity in the programme. This has been a recurring dream in the UK (Yorke, 2008) but efforts to do so have been largely centred on the US (McGrath et al., 2015).

Part of the reason for this is that the nature of US higher education makes it more realistic to search for broad agreement about which learning outcomes are most important. Firstly, the widespread focus on general education in undergraduate programmes generates consensus about learning outcomes. For example, Arum and Roksa (2011) justify their use of the CLA in their influential study on the plausible grounds that there is a common acceptance among US institutions about the importance of general critical thinking and related general skills, reflected in periodic calls for comparative student outcome measures to be used in the accreditation process (Ewell, 2015). Secondly, there are well-defined groups of institutions who broadly agree about key learning outcomes. The liberal arts colleges are the best example of this, having an explicit focus on a broad-based education and the development of general attributes such as written and oral communication, critical thinking and ethical reasoning (Association of American Colleges and Universities, 2005). These common goals of liberal arts institutions allowed the Wabash study to meaningfully administer a range of instruments assessing students' general skills, including critical thinking and moral reasoning (Pascarella & Blaich, 2013).

However, unlike the US, English higher education does not have an explicit focus on general education. Students may take a small number of broader 'elective' classes, but nearly all of their time will be spent studying within a relatively narrow field (or two narrow fields, in the case of joint programmes). For example, students at Harvard are currently only required to take 56 of 128 credits in their subject specialism over the four years of their degree (Harvard University, 2019). Most English students studying for single honours can have all of their credits in their subject specialism over the three years of their degree. Similarly, in the UK students almost always enter university on a programme with a specified subject specialism, whereas in the US students specify their specialisation after only one or two years of study. There is also a relatively high degree of specialisation in the English school system, with students typically leaving with qualifications in only three subjects. In the US, by contrast, has a broad-based secondary school curriculum and college entry is normally based on a student's SAT score, which measures general mathematical, reading and writing skills.

The development and assurance of learning outcomes in the UK are in line with this level of relative specialisation, as they are undertaken by the discipline communities themselves. The primary way of ensuring that institutions are assessing students in the 'right' way (both in terms of content and standard) is the external examining system, which is a process of peer-review in-

ternal to the discipline, largely devoid of comparable learning gain metrics. Professional disciplines often need to satisfy requirements placed on them by their professional bodies; again, this process is internal to the discipline. Non-disciplinary processes for determining and assuring learning outcomes—at institutional- or sector-level—are standardly at a very high level and are generally limited to checks that the appropriate discipline-level quality processes have been adhered to. Subject benchmarks, which are broad descriptions of what students should learn in a particular discipline, play a sector-level role and are owned by a sector-level body—the Quality Assurance Agency—but they are developed by representatives of the disciplinary communities. In England therefore, it is true to say that the system of checks and balances around the undergraduate curriculum assumes that the ultimate arbiters of what students should learn in their time in higher education are the disciplinary communities. Non-disciplinary agents (institutions, government and non-disciplinary sector bodies) have limited influence over learning outcomes, which is generally limited to ensuring that the relevant within-discipline processes have been followed.

Given the emphasis on discipline specialisation in England, efforts to mimic US developments of generic learning gain instruments are ambitious at best, and potentially misguided. In addition to differing structures of degrees in the two countries, the US efforts to measure learning gain were addressing different issues than the UK. Subsequently, 'what' was being, 'why' it was being measured and 'how' it was measured do not allow for straightforward policy transfer. However, political interest in a generic instrument led the UK to attempt to use the same methods as in the US, without thinking about the rationale underpinning the design and use of the metrics.

5. Disciplinary Learning in National Contexts

Despite a policy impetus, there are therefore a number of formidable obstacles to the development and use of generic instruments to measure learning gain in England. For example, general skills would need to be assessed in a generic instrument when students have learnt those skills almost entirely in disciplinary contexts. Even in the US with its traditional focus on general education, there is evidence that students' performance on a generic instrument such as the CLA is influenced by their field of study (Arum & Roska, 2008). The explosive impact of the 2011 study by Arum and Roska was based partly on the finding that students from fields that do not emphasise reading and writing perform less well on the CLA. This is unsurprising: with the best will in the world, the challenge of devising a test of general skills that does not discriminate between a history student and a physics student is daunting.

However, the deeper challenge concerns the authority to decide what the key learning outcomes of higher education are. The high-stakes measurement of learning gain requires fundamental decisions about what students are expected to learn. Very little in the structures of English higher education indicate that that is appropriate for non-disciplinary agents—government, regulator, funding body, quality agency—to make those determinations. As described above, English higher education treats disciplinary academic communities as the ultimate arbiters of what students should learn. This does not rule out the development of generic instruments to measure learning gain. A disciplinary community may decide that general skills (e.g., numerical reasoning) are among their important learning outcomes, and that those skills can be validly assessed using generic assessment tools. However, the structures of English higher education indicate that the decision would rest with the disciplinary community; no non-disciplinary agent could persuasively claim the authority to decide what students ought to learn.

The recent developments in the measurement of learning suggest the role of the disciplines in determining and assuring what students should be learning is under question. The attempt by sector-wide, non-disciplinary agents to create instruments to measure learning gain, and by doing so to implicitly claim authority over the key learning outcomes of higher education, fits with broader patterns of administrative and managerial encroachment on academic authority: 1) the more assertive behaviour of administrative agents (Bleiklie, 1998); 2) the more hands-on role of management (Deem, 2017), the usurpation of professional expertise by management expertise (Amaral, Meek, Larsen, & Lars, 2003) inspired by the reduction in trust in professional expertise (Beck & Young, 2005); and 3) the demystification of academic work in order to facilitate its management using generic tools and techniques (Henkel, 1997). The literature on managerialism in higher education focuses on the increasingly muscular presence of administrative and managerial units within institutions, but a parallel process has been occurring at sector-level, with organisations such as the Quality Assurance Agency and the Office for Students taking on increasing power within themselves at the expense of disciplinary communities (Becher & Trowler, 2001; Filippakou & Tapper, 2019). The amplification of the market in the English higher education system—increased fees, removal of number caps, introduction of 'kitemarks' via the judgements of the Teaching Excellence Framework—has coincided with encroachments on the responsibilities of academics, such as frequent accusations by (successive) higher education Ministers that they are failing to maintain appropriate standards and allowing 'grade inflation.'

6. Learning Gain and the Purpose of Higher Education

The attempt to quantify learning raises questions about the purpose and underpinning values of higher education and necessitates debate about the rationale for

quantification—whether it is for accountability, measuring performance, assuring quality or for the enhancement of teaching, learning and the student experience. Metrics have many uses, but there is inherent tension between metrics used for accountability and improvement (Kuh & Ewell, 2010). Through focusing on 'how' to measure learning gain, the learning gain programme of work did not address the question of what quality is in higher education, or the more profound question of what higher education is for; the answers have a significant impact on the use of any resulting data. There is a 'paradoxical tension' between how academic staff and external stakeholders view accountability by student learning outcomes (Borden & Peters, 2014). The assumption that it is in the gift of government and sector-level funding bodies and regulators to define measures of learning gain usurps the authority of disciplines as the arbiters of student learning. The absence of student voices also raises questions about their role in determining what their educational experience is for (Klemenčič, 2018).

In terms of assuring quality, there has been a broad shift from process and programme evaluation to outcome evaluation (Harvey & Williams, 2010). For example, there is increasing emphasis on salary data (drawing on the Longitudinal Education Outcomes dataset) as a metric of educational quality (Office for Students, 2019a). When it comes to learning gain, the tension around who 'owns' the measures has implications for evaluating performance. As found across the pilot projects, disciplinary differences in marking present challenges of using outcome data for cross-subject and institutional comparisons (Ylonen, Gillespie, & Green, 2018). Sector bodies such as funding councils and the new regulator work at institutional level. However, unless metrics have resonance at the disciplinary level, where students experience higher education, they will fail to meet the ultimate aims of assuring and improving the experience of students, in addition to lacking the legitimacy conferred by disciplinary authority. Desire for comparable metrics leads to a focus on standardized outcome tests over instruments designed to support student learning and enhance teaching (Douglass, Thomson, & Zhao, 2012).

7. Quantification as an End in Itself

The search for comparable information about student learning has led to a focus on the 'quantity' of learning a student receives from their investment in higher education. This simplistic quantification of learning ignores the merit of the content and the process of learning. Any measure of learning gain would always be a proxy of the activity itself; however, without a clear purpose for measuring and quantifying learning the proxy measures become divorced from the underlying activity. Furthermore, through using proxy measures in high-stakes quality frameworks, they become targets in themselves. This has been seen through the use of proportion of top grades awarded in league tables, and the recent rapid escalation in grades across the UK sector (Palfreyman, 2019). Similarly, in the US the use of admission rate and yield metrics (the ratio of admitted students and those that matriculate) have dramatically impacted admissions practices in the US (Monks & Ehrenberg, 1999).

A lack of a rationale, beyond the initial ministerial catalyst, for measuring learning gain beset the learning gain programme. In the pilot projects, academics worried about 'unintended' use of metrics or 'non-disclosed intentions' around their use. Several projects concluded they would rather err on the side of not producing national measures rather than developing them and then hoping they were used for 'good' educational purposes. When learning gain is separated from debates about purpose, it allows available numbers to be used as proxy measures, resulting in many higher education metrics that are divorced from causal effects of institutions (Matsudaira, 2016).

There are wide ranging consequences of using proxy measures, particularly for vulnerable and disadvantaged groups (O'Neil, 2017), such as through geographical measures of deprivation that ignore individual circumstances and algorithms that normalise explained and unexplained attainment gaps by ethnicity (Office for Students, 2019b, 2019c). Social inequalities are perpetuated through quality judgements based on institutional reputation, a key sorting and selection criterion for many employers (Hazelkorn, 2015). In response many employers now design in-house recruitment mechanisms. These are often methodologically flawed and burdensome tests, which creates high inefficiencies for employers and graduates (Keep & James, 2010). Furthermore, numbers as proxies become ends in themselves:

> The net result is that ranks become naturalised, normalised and validated, through familiarity and ubiquitous citation, particularly through recitation as 'facts' in the media. Rankings, thus, attain an unwarranted truth status that makes them self-fulfilling by virtue of their persistence and existence. (Lynch, 2015, p. 198)

The quantification of learning can distil a complex activity to a number, but without a rationale for developing, selecting and using measures the number loses any sense of purpose or meaning and becomes an end in itself. Learning gain becomes another metric to be used for marketing purposes (Polkinghorne, Roushan, & Taylor, 2017). Additionally, as a data-driven regulator, the Office for Students has also set key performance indicators for itself, with a measure of learning gain being one its 26 'Measures of Success' (Office for Students, 2019d), meaning that the regulator needs to develop a measure *for its own use*.

Despite the challenges described in this article, the measurement of learning gain has immense potential for enhancing quality and performance in higher education (Kuh & Jankowski, 2018; Shavelson et al., 2018). For

example, developing 'quantity' measures of quality facilitates policy drives for competition, transparency and accountability, which are unlikely to dissipate. In the search for valid measures of teaching quality, learning gain—particularly when used as the basis for calculating the 'value added' by institutions and programmes—has benefits over proxy metrics such as student satisfaction and salary data. Quantification approaches could also in principle help align various disciplinary-based quality approaches, addressing concerns around equity of experience and differential outcomes (Kandiko Howson & Mawer, 2013). However, through focusing on 'how' to measure learning gain independent of 'why' to measure it, or 'what' to measure, the creation of a robust higher education quality system with comparable student outcomes and clear evidence of value for money has been set back by these recent developments. With a quality system aligned to disciplines, yet a regulatory system that holds institutions to account, simple, straightforward measures of the quality of what students are gaining in higher education have not emerged. As long as the disciplines act as the arbiters of quality in education, a debateable position itself, the development of meaningful institutional-level measures will be challenging.

8. Conclusion

The search for data about learning gain provides an illustrative example of the 'evaluative state' in English higher education. Sector agencies engage in efforts to develop quantitative instruments in areas where they have no explicit claim to authority, relying on a general sense of the right of administrative and managerial agents to monitor the outcomes of higher education institutions. Logics inherent elsewhere in the system—about the awesome technical challenges in measuring learning gain across disciplines and institutions, about the unintended impact of quality metrics, about the tension between accountability and improvement, about the lack of apparent purchase that quantitative indicators of teaching quality have on student recruitment, about the role of disciplines in determining and assuring learning outcomes—are overridden by the quantitative rationale. Developments that assume particular answers to fundamental questions about the value of higher education take place without any explicit consideration of those questions. The answers are provided by the systems and structures that have particular perspectives—managerialism, quantification—built in. Higher education is full of contentious developments that adopt the logic of quantification without explicit discussion and undermine or usurp traditional disciplinary-based methods of quality assurance, accountability and regulation. The search for sector-wide measures of learning gain in English higher education provides a limit to governance by numbers, and an example of the overextension of the logic of quantification and a failure to turn 'what' students learn into 'how much' was gained.

Acknowledgments

We would like to thank Maarten Hillebrandt and Michael Huber for organizing the workshop and thematic issue and for the helpful feedback from three anonymous reviewers.

Conflict of Interests

The authors declare no conflict of interests.

References

Amaral, A., Meek, V. L., Larsen, I. M., & Lars, W. (Eds.). (2003). *The higher education managerial revolution?* (Vol. 3). Berlin and Heidelberg: Springer Science + Business Media.

Arum, R., & Roska, J. (2008). *Learning to reason and communicate in college: Initial report of findings from the CLA longitudinal study*. New York, NY: Social Science Research Council.

Arum, R., & Roska, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Association of American Colleges and Universities. (2005). *Liberal education outcomes: A preliminary report on student achievement in college*. Washington, DC: Association of American Colleges and Universities.

Bachan, R. (2017). Grade inflation in UK higher education. *Studies in Higher Education*, *42*(8), 1580–1600.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Buckingham: Open University Press.

Beck, J., & Young, M. F. (2005). The assault on the professions and the restructuring of academic and professional identities: A Bernsteinian analysis. *British Journal of Sociology of Education*, *26*(2), 183–197.

Birkland, T. A. (2015). *An introduction to the policy process: Theories, concepts, and models of public policy making*. Abingdon: Routledge.

Bleiklie, I. (1998). Justifying the evaluative state: New public management ideals in higher education. *Journal of Public Affairs Education*, *4*(2), 87–100.

Borden, V. M., & Peters, S. (2014). Faculty engagement in learning outcomes assessment. In H. Coates (Ed.), *Higher education learning outcomes assessment: International perspectives* (pp. 201–212). Bern: Peter Lang GmbH.

Browne, J. (2010). *Securing a sustainable future for higher education: An independent review of higher education funding and student finance* (Report BIS/10/1208). London: Department for Business, Innovation and Skills.

Center of Inquiry. (2016). Wabash national study 2006–2012. *Centre of Inquiry*. Retrieved from https://centerofinquiry.org/wabash-national-study-of-liberal-arts-education

Coates, H., & Mahat, M. (2014). Advancing student learning outcomes. In H. Coates (Ed.), *Higher education learning outcomes assessment: International perspectives* (pp. 15–31). Bern: Peter Lang GmbH.

Coates, H., & McCormick, A. (Eds.). (2014). *Engaging university students: International insights from system-wide studies*. London: Springer.

Cranmer, S. (2006). Enhancing graduate employability: Best intentions and mixed outcomes. *Studies in Higher Education*, *31*(2), 169–184.

Deem, R. (2017). New managerialism in higher education. In J. C. Shin & P. Teixeira (Eds.), *Encyclopaedia of international higher education systems and institutions* (pp. 1–7). Dordrecht: Springer.

Department for Business, Innovation and Skills. (2011). *Higher education: Students at the heart of the system*. London: Department for Business, Innovation and Skills.

Department for Business, Innovation and Skills. (2016). *Success as a knowledge economy: Teaching excellence, social mobility & student choice*. London: Department for Business, Innovation and Skills.

Douglass, J. A., Thomson, G., & Zhao, C. M. (2012). The learning outcomes race: The value of self-reported gains in large research universities. *Higher Education*, *64*(3), 317–335.

Ehrenberg, R. G. (2003). Reaching for the brass ring: The US News & World Report rankings and competition. *The Review of Higher Education*, *26*(2), 145–162.

Ewell, P. (2015). *Transforming institutional accreditation in US higher education*. Boulder, CO: National Center for Higher Education Management.

Excellence in Research for Australia. (2018). Australian government. *Australian Research Council*. Retrieved from https://www.arc.gov.au/excellence-research-australia/era-2018

Filippakou, O., & Tapper, T. (2019). The state, the market and the changing governance of higher education in England: From the University Grants Committee to the Office for Students. In O. Filippakou & T. Tapper (Eds.), *Creating the future? The 1960s new English universities* (pp. 111–121). Cham: Springer.

Frankham, J. (2016). Employability and higher education: The follies of the 'productivity challenge' in the Teaching Excellence Framework. *Journal of Education Policy*, *32*(5), 628–641.

Giroux, H. (2002). Neoliberalism, corporate culture, and the promise of higher education: The university as a democratic public sphere. *Harvard Educational Review*, *72*(4), 425–464.

Gunn, A. (2018). The UK Teaching Excellence Framework (TEF): The development of a new transparency tool. In A. Curaj L. & R. Pricopie (Eds.), *European higher education area: The impact of past and future policies* (pp. 505–526). Cham: Springer.

Hart Research Associates. (2015). *Falling short? College learning and career success*. Washington, DC: Association of American Colleges and Universities.

Harvard University. (2019). Harvard University handbook for students 2019–2020. *Harvard University*. Retrieved from https://handbook.fas.harvard.edu/book/welcome

Harvey, L., & Williams, J. (2010). Fifteen years of quality in higher education. *Quality in Higher Education*, *16*(1), 3–36.

Hazelkorn, E. (2015). *Rankings and the reshaping of higher education: The battle for world-class excellence*. Cham: Springer.

Henkel, M. (1997). Academic values and the university as corporate enterprise. *Higher Education Quarterly*, *51*(2), 134–143.

Higher Education Funding Council for England. (2018). Learning gain. *Higher Education Funding Council for England.* Retrieved from https://webarchive.nationalarchives.gov.uk/20180319113650/http://www.hefce.ac.uk/lt/lg

Kandiko Howson, C. B. (2019). *Final evaluation of the Office for Students learning gain pilot projects*. Bristol: Office for Students.

Kandiko Howson, C. B., & Buckley, A. (2017). Development of the UK engagement survey. *Assessment & Evaluation in Higher Education*, *42*(7), 1132–1144.

Kandiko Howson, C. B., & Mawer, M. (2013). *Student expectations and perceptions of higher education*. London: King's College London.

Kandiko Howson, C. B., & Weyers, M. (Eds.). (2013). *The global student experience: An international and comparative analysis*. London: Routledge.

Keep, E., & James, S. (2010). *Recruitment and selection: The great neglected topic* (SKOPE Research Paper 88). Cardiff: Cardiff University and SKOPE.

Klemenčič, M. (2018). The student voice in quality assessment and improvement. In E. Hazelkorn, H. Coates, & A. C. McCormick, (Eds.), *Research handbook on quality, performance and accountability in higher education* (pp. 332–346). Cheltenham: Edward Elgar Publishing.

Kuh, G. D., & Ewell, P. T. (2010). The state of learning outcomes assessment in the United States. *Higher Education Management and Policy*, *22*(1), 1–20.

Kuh, G. D., & Jankowski, N. A. (2018). Assuring high-quality learning for all students: Lessons from the field. In E. Hazelkorn, H. Coates, & A. C. McCormick (Eds.), *Research handbook on quality, performance and accountability in higher education* (pp. 305–320). Cheltenham: Edward Elgar Publishing.

Lynch, K. (2015). Control by numbers: New managerialism and ranking in higher education. *Critical Studies in Education*, *56*(2), 190–207.

Matsudaira, J. (2016). Defining and measuring institutional quality in higher education. In K. Matchett, M. Lund Dahlberg, & T. Rudin (Eds.), *Quality in the undergraduate experience: What is it? How is it measured? Who decides?* (pp. 57–80). Washington, DC: National Academies Press.

McGrath, C. H., Guerin, B., Harte, E., Frearson, M., &

Manville, C. (2015). *Learning gain in HE*. Cambridge: Rand Corporation.

Monks, J., & Ehrenberg, R. G. (1999). US News & World Report's college rankings: Why they do matter. *Change: The Magazine of Higher Learning*, *31*(6), 42–51.

Naidoo, R., & Williams, J. (2015). The neoliberal regime in English higher education: Charters, consumers and the erosion of the public good. *Critical Studies in Education*, *56*(2), 208–223.

O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, NY: Crown.

Office for Students. (2019a). Graduate earnings data on Unistats from the Longitudinal Education Outcomes (LEO) data. *Office for Students*. Retrieved from https://www.officeforstudents.org.uk/data-and-analysis/graduate-earnings-data-on-unistats

Office for Students. (2019b). Ethnicity. *Office for Students*. Retrieved from https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/evaluation-and-effective-practice/ethnicity

Office for Students. (2019c). Young participation by area. *Office for Students* Retrieved from https://www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area

Office for Students. (2019d). Measures of our Success. *Office for Students* Retrieved from https://www.officeforstudents.org.uk/about/measures-of-our-success

Olssen, M. (2016). Neoliberal competition in higher education today: Research, accountability and impact. *British Journal of Sociology of Education*, *37*(1), 129–148.

Organisation for Economic Co-operation and Development. (2013a). *Assessment of higher education learning outcomes* (Feasibility Study Report, Volume 2). Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. (2013b). *Assessment of higher education learning outcomes* (Feasibility Study Report, Volume 3). Paris: Organisation for Economic Co-operation and Development.

Palfreyman, D. (2019). Regulating higher education markets. In T. Strike, J. Nicholls, & J. Ruthforth (Eds.), *Governing higher education today: International perspectives* (pp. 202–216). London: Routledge.

Pascarella, E., & Blaich, C. (2013). Lessons from the Wabash national study of liberal arts education. *Change*, *45*(2), 6–15.

Polkinghorne, M., Roushan, G., & Taylor, J. (2017). Considering the marketing of higher education: The role of student learning gain as a potential indicator of teaching quality. *Journal of Marketing for Higher Education*, *27*(2), 213–232.

Power, M. (1994). *The audit explosion*. London: Demos.

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior*, *60*, 333–341.

Shavelson, R. J., Zlatkin-Troitschanskaia, O., & Mariño, J. P. (2018). Performance indicators of learning in higher education institutions: An overview of the field. In E. Hazelkorn, H. Coates, & A. C. McCormick (Eds.), *Research handbook on quality, performance and accountability in higher education* (pp. 249–263). Cheltenham: Edward Elgar.

Tymon, A. (2013). The student perspective on employability. *Studies in Higher Education*, *38*(6), 841–856.

Ylonen, A., Gillespie, H., & Green, A. (2018). Disciplinary differences and other variations in assessment cultures in higher education: Exploring variability and inconsistencies in one university in England. *Assessment & Evaluation in Higher Education*, *43*(6), 1009–1017.

Yorke, M. (2008). *Grading student achievement in higher education: Signals and shortcomings*. Abingdon: Routledge.

About the Authors

Camille Kandiko Howson is Associate Professor of Education at the Centre for Higher Education Research and Scholarship at Imperial College London. She is an international expert in higher education research with a focus on student engagement; student outcomes and learning gain; quality, performance and accountability; and gender and prestige in academic work. She is a Principle Fellow of the Higher Education Academy.

Alex Buckley is an Assistant Professor in the Learning and Teaching Academy at Heriot Watt University in Edinburgh, Scotland. His work is focused on supporting individuals and groups to enhance learning and teaching. He has previously held roles at Strathclyde University and the Higher Education Academy (now AdvanceHE). His research interests include assessment and feedback, student engagement and student surveys.

Politics and Governance (ISSN: 2183–2463)
2020, Volume 8, Issue 2, Pages 15–25
DOI: 10.17645/pag.v8i2.2594

Article

Algorithmic Allocation: Untangling Rival Considerations of Fairness in Research Management

Guus Dix [1,*], Wolfgang Kaltenbrunner [1], Joeri Tijdink [2,3], Govert Valkenburg [4] and Sarah de Rijcke [1]

[1] Centre for Science and Technology Studies, Leiden University, 2311 EZ Leiden, The Netherlands;
E-Mails: g.dix@cwts.leidenuniv.nl (G.D.), w.kaltenbrunner@cwts.leidenuniv.nl (W.K.),
s.de.rijcke@cwts.leidenuniv.nl (S.d.R.)
[2] Department of Medical Humanities, AmsterdamUMC, location VUmc, 1081 HV Amsterdam, The Netherlands;
E-Mail: jk.tijdink@amsterdamumc.nl
[3] Department of Philosophy, Vrije Universiteit Amsterdam, 1081 HV Amsterdam; The Netherlands
[4] Department of Interdisciplinary Studies of Culture, Norwegian University of Science and Technology, 7491 Trondheim, Norway; E-Mail: govert.valkenburg@ntnu.no

* Corresponding author

Submitted: 30 October 2019 | Accepted: 30 December 2019 | Published: 9 April 2020

Abstract
Marketization and quantification have become ingrained in academia over the past few decades. The trust in numbers and incentives has led to a proliferation of devices that individualize, induce, benchmark, and rank academic performance. As an instantiation of that trend, this article focuses on the establishment and contestation of 'algorithmic allocation' at a Dutch university medical centre. Algorithmic allocation is a form of data-driven automated reasoning that enables university administrators to calculate the overall research budget of a department without engaging in a detailed qualitative assessment of the current content and future potential of its research activities. It consists of a range of quantitative performance indicators covering scientific publications, peer recognition, PhD supervision, and grant acquisition. Drawing on semi-structured interviews, focus groups, and document analysis, we contrast the attempt to build a rationale for algorithmic allocation—citing unfair advantage, competitive achievement, incentives, and exchange—with the attempt to challenge that rationale based on existing epistemic differences between departments. From the specifics of the case, we extrapolate to considerations of epistemic and market fairness that might equally be at stake in other attempts to govern the production of scientific knowledge in a quantitative and market-oriented way.

Keywords
algorithmic allocation; higher education; marketization; performance indicators; quantification; resource allocation

Issue
This article is part of the issue "Quantifying Higher Education: Governing Universities and Academics by Numbers" edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

1. Introduction

Large research organizations need to address a question that is at once moral and technical: How to allocate resources across different departments to effectively foster certain organizational goals and be perceived as generally fair at the very same time. The organization central to this article—a university medical centre—sought the answer to this question in an allocation model based on quantitative performance criteria and departmental competition. More specifically, the Centre intended to 'stimulate the control of the (financial) business operation through the implementation of performance-based financing at the departmental level' and to gauge excellence in research 'based on objective and measurable criteria' such as 'the citation score and impact factor' (doc-

ument 1; document 5). In 2015 and 2016, the board of directors was engaged in an iterative process of developing an allocation algorithm together with financial controllers and (bio)medical scientists. Algorithmic allocation, as we call it, is a form of quantitative reasoning that makes it possible to calculate the costs and the achievements of a department without engaging in a detailed qualitative assessment of the current content and future potential of its research activities. This form of reasoning did not come out of thin air. In fact, it fits into a history of attempts to induce departments to increase their scientific performance which is underpinned by entrenched ideas of fairness and effectiveness in research management. Entrenched as these ideas might be, the board decided to put the model on hold only two years after it had been introduced. In that short timeframe, algorithmic allocation had become controversial for the way it affected the research capacity of different departments. As one interviewee put it succinctly: 'I think in fact that you do not have a backbone, sort of, when you say: "Yes, we will let the algorithm decide where the money will go to"' (professor 2).

Algorithmic allocation is tied to a particular organizational setting at a particular time and simultaneously linked to more generic developments in higher education—and society at large—such as quantification, marketization, and algorithm-based decision making. Over the past four decades, quantitative performance indicators and market-oriented ideals and practices have become more significant in academia. University rankings, for instance, have re-emerged at the beginning of this millennium at the intersection of private initiatives and a new field of research analytics and secured themselves a strong position in university management (Hammarfelt, de Rijcke, & Wouters, 2017; Paradeise & Filliatreau, 2016; Paradeise & Thoenig, 2018; Sauder & Espeland, 2009). Beyond managerial contexts, quantitative indicators such as the Journal Impact Factor—as a proxy for the quality of scientific journals—and the H-index—as a proxy for the quality of a scientist—have also consolidated themselves on the shop floor of academic research (Rushforth & de Rijcke, 2015, 2017; Rushforth, Franssen, & de Rijcke, 2018). The increased emphasis on competition, incentivization, and the economic value of research has led social scientists to speak of 'market universities' and 'epistemic capitalism' as the new academic status quo (Berman, 2012; Birch, 2017; Fochler, 2016; Slaughter & Rhoades, 2009). Sociologists and new media studies scholars have further explored the ways in which marketization and quantification are tied to increasing automation and algorithmic calculation (Berman & Hirschman, 2018; Couldry, 2016; Finn, 2017; van Dijck, 2013). In an 'algorithmic culture' (Striphas, 2015), human thoughts and conducts are easily expressed in the logic of big data. These computational practices are further embedded in highly consequential decision-making processes in a wide range of domains: Who should obtain a visa? Who should be considered a risk to society? Who deserves to be rewarded? And which product or service will satisfy our desires? When value is increasingly measured computationally and algorithms inform—or make—decisions, it is imperative to better understand the way algorithms are designed and how they shape public spheres.

This article focuses on higher education as one of the key areas of social life where market-oriented practices, quantitative performance indicators, and algorithm-based decision-making processes have made large inroads. More particularly, it extends the literature discussed above in the direction of the normative considerations that people offer to support or problematize these developments. For whatever their actual consequences, such developments do not take place without attempts being made to support or challenge them. The research question of our article is analytical: Which normative considerations do researchers and administrators draw upon to justify or criticize algorithm-based decision making in academia? We provide an answer to this question based on focus groups, semi-structured interviews, and document analysis. More specifically, we untangle two rival notions of fairness at play in research management: Considerations of 'market fairness' are offered in support of algorithmic allocation whereas considerations of 'epistemic fairness' are offered as a challenge to it (Section 2). Our analysis starts with an overview of the current state of algorithmic allocation at the university medical centre (Section 3). Subsequently, we analyze the construction of a market-oriented justification that merges a critique of 'unfair' privileges with an emphasis on 'fair' competition, incentives, and exchange (Section 4). We conclude our analysis with attempts to challenge algorithmic allocation by those who stress the unfairness of a uniform model given the significant epistemic differences between departments (Section 5). Some people might be more inclined to favour consideration of market fairness over considerations of epistemic fairness (and vice versa). But it is important to keep in mind that the central tension between the two is not only found *between* relatively fixed sets of actors but also *within* actors themselves. Based on our analytic distinction, we caution against having a narrow perspective of research organizations as being relatively homogenous entities and mere recipients of broader trends in academia, and furthermore, we offer a set of questions to enable researchers and research managers to think reflexively about algorithm-based decision-making (Section 6).

2. Conceptualizing Market Fairness and Epistemic Fairness: Theory and Method

Markets are important economic phenomena but they are equally important for the moral justifications they enable. Sociologists have theorized how markets are linked to particular 'moral views' on what binds people together (Boltanski & Thevenot, 2006; Fourcade & Healy,

2007). Four main aspects of this 'market bond' have been highlighted. Firstly, Boltanski and Thevenot (2006, pp. 196–197) speak of it as a relatively coherent set of normative principles which place much emphasis on 'competition between beings placed in a state of rivalry' and on the value of competitive achievement. Second, to get ahead in a competitive environment, people lend themselves 'willingly to every opportunity to engage in a transaction' and should thus be enabled to exchange goods and services as they see fit (Boltanski & Thevenot, 2006, p. 200). The third aspect of the market bond is that competition and exchange should induce people to work for the benefits of others—not just for themselves. In that sense, 'market systems are supposed to provide incentives and opportunities for innovation' (Fourcade & Healy, 2007, p. 290). Fourth, markets are considered to be a fair allocation system vis-à-vis a system of government interventions that privilege some at the cost of others. Thinking along these lines, 'competitive economic arrangements are the best defence…against arbitrary interference' (Fourcade & Healy, 2007, p. 290). In our case, 'market fairness' is the analytical term that captures these four normative considerations—competitive achievement, exchange, incentivization, and limiting arbitrary interference—with which researchers and administrators build an organizational rationale to support algorithmic allocation.

Considerations of market fairness are important but not the sole source of normativity. As we briefly discussed, algorithmic allocation quickly became controversial for its practical consequences and its normative underpinnings. Researchers and administrators highlight major differences in the way scientific knowledge is produced in different departments to dispute the alleged fairness of a uniform allocation model. The idea that practices of knowledge production vary greatly is not new to scholars in Science and Technology Studies, an interdisciplinary field of social scientists who study the social, material, and cognitive aspects of scientific inquiry and technological developments. Knorr-Cetina (1999), for instance, has coined the term 'epistemic cultures' to capture differences in the environment, procedures, and objects of different research communities as well as the different affordances and constraints these communities experience (see also Borgman, 2017; Franssen, Scholten, Hessels, & de Rijcke, 2018; Hessels, Franssen, Scholten, & de Rijcke, 2019). Following Knorr-Cetina, we can say that a university medical centre is unified in its focus on the understanding, prevention, and treatment of illness but internally divided in terms of the underlying epistemic cultures. The considerations of 'epistemic fairness' we discerned call algorithmic allocation into question by highlighting crucial differences in material costs, technological requirements, access to resources, and collaboration and publication practices.

To analyse rival considerations of fairness in research management, we combined semi-structured interviews with focus groups and document analysis. The first, exploratory research phase of our research project consisted of two focus groups with six early career and eight senior (bio)medical researchers, one interview with the dean and one interview with three research policy advisors. The interviews and focus groups initially zoomed in on ideals of responsible (bio)medical research and on instruments to foster innovative and societally relevant research practices. The analysis of the interview transcripts led to the selection of three main areas of organizational change at the centre: reorganizing research, measuring and visualizing scientific output, and evaluating and rewarding scientists. These three themes were central to the second, more focused research phase. This second phase consisted of eight additional interviews with researchers as well as staff members in key advisory positions within the Centre: two research policy advisors, two financial controllers, and four professors involved in central management. In addition to the analysis of interview transcripts and notes, the article is based on a series of documents that capture different features of research organization, measurement, and evaluation at the Centre: The annual reports on the financial and organizational state of the Centre over an eleven-year period; two five-year plans in which the Centre gives a broad outline of the main developments and strategies in (bio)medical research and healthcare practices; the hiring and promotion procedures for associates and full professors; website material; a dissertation on early developments in resource allocation (not in the bibliography for reasons of anonymity); documents about the new resource allocation model; as well as internal presentations about the model before it was put into practice (see Table 1 for a detailed overview of the interviews and focus groups and the selection of documents used for this study).

3. Algorithmic Allocation: Towards a Metric-Based Model

In a very basic form, algorithmic allocation goes back to the very establishment of the medical centre during a wave of mergers between academic hospitals and university-based medical faculties in the late 1990s and early 2000s. At the time, the board of directors decided to put a small part of its research budget in a separate fund from which it was subsequently distributed back to the departments based on two rather straightforward performance criteria: the number of doctorates and the number of 'full-time equivalents' acquired through external funding. That allocation model remained more or less unchanged for over a decade (document 6). Only recently, in 2015, did the board decide to shift from 'performance-based funding' to 'performance-based financing.' More than a mere semantic change, the intended reform was substantial: '[S]o we moved on by allotting all gains and all costs to the department. And we made a part of it variable, a pretty large part of it at that' (financial controller 1). Two substantial changes stand out.

First, performance-based financing went far beyond performance-based funding in making departments responsible for the costs they made in conducting their research. In terms of overhead costs, departments were now required to pay the central administration for the use of office and laboratory space—in square meters—and for the number of their employees (as a parameter to divide all overhead costs). The integration of such costs into the allocation model was not primarily driven by developments inside the field of (bio)medical research but came primarily from healthcare (financial controller 2). An earlier political push for the marketization of Dutch healthcare had led hospitals to become more sensitive to the costs and the rewards of medical interventions. Putting a price on the use of medical facilities and personnel for every intervention was a way to increase cost-sensitivity. And as the centre housed both medical and research facilities the question about the overhead costs arose there too: '[S]o at a certain moment, it was buzzing around, like: We have to do performance-based financing because they are doing it too' (financial controller 1).

Second, performance-based financing went beyond performance-based funding in doubling the amount of money that was distributed based on research results. The fund to reimburse departments for their performance was split into two equal parts—one relatively fixed, the other more variable. The first half of the budget was allocated based on the number of associate and full professors employed at the department compared to the total number of full-time jobs at the medical centre. The second half was variable and allocated based on quantitative performance indicators. The range of potential indicators was restricted: 'There are not that many parameters available in the world of research that can also be measured well, to put it that way. So, then…you soon fall back upon the happy few' (financial controller 2). After many rounds of consultation with researchers and financial controllers, the board decided upon the following four performance indicators in 2016: 'acquisition power,' 'doctorates,' 'author force,' and 'score top publications' (see Table 1 for an overview). 40 percent of the variable research budget was reserved for 'acquisition power' and 20 percent for the other three indicators. The overall score of a department vis-à-vis other departments determined its so-called 'market share' in the budget (document 2). In addition to the algorithmic allocation model, these performance indicators were also used to flag the position of the centre in the international research landscape and included in the hiring and promotion guidelines for assistant, associate, and full professors (document 7, 8 and 9).

The first two indicators are concerned with the financial and human resources that departments themselves are able to attract. The indicator 'acquisition power' covers the success of a department in bidding for external grants. The ability to attract such grants has become important in the Dutch academic system over the past

Table 1. Overview of interviews, focus groups, and documents.

Interviews and focus groups	Participants	Date
Interview 1	Three advisors research policy	December 27, 2017
Interview 2	Dean	February 8, 2018
Focus Group 1	Six early career researchers (four PhD students, two postdoctoral researchers)	June 18, 2018
Focus Group 2	Eight senior researchers (two assistant professors, four associate professors, and two full professors)	June 18 2018
Interview 3	Advisor research policy 1	December 11, 2018
Interview 4	Advsisor research policy 2	December 11, 2018
Interview 5	Professor 1	March 13, 2019
Interview 6	Financial controller 1	March 19, 2019
Interview 7	Professor 2	March 19, 2019
Interview 8	Financial controller 2	March 27, 2019
Interview 9	Professor 3	April 19, 2019
Interview 10	Professor 4	May 16, 2019
Documents	**Brief description**	
Document 1	Annual report 2017	
Document 2	Internal memo on allocation model, 2018	
Document 3	Internal presentation on allocation model, 2015	
Document 4	Internal presentation on performance indicators, 2016	
Document 5	Strategic five-year plan, 2018	
Document 6	Dissertation on early developments in resource allocation, 2009	
Document 7	Hiring and promotion guideline for assistant and associate professor	
Document 8	Hiring and promotion guideline for full professor	
Document 9	Online presentation of the center	

two decades as universities became more dependent on the ability of individual researchers to successfully attract the resources—and the ensuing individual and institutional prestige—to carry out larger research projects. The acquisition power of a department is calculated by taking the three-year average of the grants acquired. The second indicator—'doctorates'—covers the relative success of a department in attracting and supervising PhD students. More than just a reward, the resources are also meant to compensate departments for the insufficient government funding for PhDs. The indicator is calculated by taking the mean number of dissertations defended at a department over the past three years. The final two cover scientific output and are somewhat more laborious. The third indicator is called 'author force' and covers the publication successes of researchers in terms of quantity and impact. The author force of a department is calculated by multiplying the number of articles published over the past three years with the Mean Normalized Citation Score (MNCS). The MNCS is a score that expresses the scientific impact of a set of publications in a particular subfield of research by comparing the citations it managed to attract with what is normal in that subfield. Because articles in (bio)medicine are usually written by a group of researchers, the articles are weighted differently according to the position of the authors. The fourth indicator is the 'score top publications' and covers publications in the highest quartile of scientific journals. The centre uses the ranking of journals by Clarivate Analytics to determine whether a publication belongs to the top one percent—valued highest in the model—or to some other percentile. Again, the department receives money based on the position of the authors on the articles in question (see Table 2).

4. In Support of Algorithmic Allocation: Market Fairness

Algorithmic allocation has been the preferred—though contested—model to distribute resources within the centre over the course of two decades. That continuity is the starting point for teasing out some of its normative underpinnings. In particular, this section analyses the place of 'market fairness' in providing a rationale to support algorithmic allocation. That concept of market fairness extends the idea that markets are not just economic phenomena out there in the world, but equally important as a reference point in the normative considerations that we draw upon to support particular ideas of merit and fairness in an organizational or political context (Boltanski & Thevenot, 2006; Fourcade & Healy, 2007). Following that line of thought, researchers and policymakers do not literally refer to the university medical centre as a marketplace but do draw upon a range of market-oriented considerations to elaborate on the fairness of this allocation model. Algorithmic allocation was justified negatively in terms of the past injustices it sought to make up for, and justified positively in terms of the principles—exchange, competition, and incentivization—that should prevail when distributing resources fairly and effectively.

In negative terms, market fairness is about eliminating unfair advantages emanating from past decisions to privilege one department at the cost of others. When people have been granted something, one professor remarks, they are not very eager to let go of it. Over time, this creates disparities between departments: '[S]o there are some historical budgets which makes it difficult sometimes….You do run into inequalities. Not everyone is always equal, I say at such moments' (professor 4). The former dean favoured algorithmic allocation because it would ameliorate this situation. As another professor who was close to the whole operation recalls:

> That's also what he explicitly said: 'It is all on historical grounds that this department gets this much money and that department that much. That is not fair. I will reconsider it all. And I will put it all in a variable fund and then we will see.' Under the guise of: 'This is fair.' (professor 2)

Table 2. Schematic overview of the allocation model (based on document 2).

Indicator	Acquisition	Dissertations	Author Force	Top Publications
Weight	40%	20%	20%	20%
Formula	funds	#doctorates	(P × A) × MNCS	(P × A) × QW
Explanation	The amount of funding acquired from research organizations and patient organizations	The number of dissertations defended at a department	'P' stands for the number of articles 'A' stands for author position: • First author: 30% • Second author: 20% • Last author: 20% • Other authors: 30% (distributed equally) 'MNCS' stands for Mean Normalized Citation Score	'P' (see previous indicator) 'A' (see previous indicator) QW' stands for a ranking of 'journals based on the Journal Impact Factor (provided by Thomson Reuters): • Top 1% of journals: 2 • Top 1–5% of journals: 1,5 • Top 5–10% of journals: 1,0 • Top 10–25%: 0,5 • Lower than top 25%: 0

Variability in the funding of departments offered a way out of inequality—but so was quantification. Originally, the historical budgets emerged from rival claims for more resources that were expressed in words, 'there are quite a lot of researchers and they all want to say to the dean: "I conduct the best research in the world." So yes, then quantitative is attractive' (professor 1). The quantification of research output offers the dean a way to stay clear of allocation decisions based on verbal claims of research excellence.

The search for quantitative performance indicators was negatively justified with reference to historical budgets but algorithmic allocation was predominantly justified in positive terms. The first positive consideration of market fairness is about the ideal of the organization as being a site of exchange. A key novelty of the allocation model was its integration of both overhead costs and research output. The internal presentation that launched the search for the best performance indicators specified that one should pay a price for the services used—the use of office space and research facilities—and be rewarded for the things one delivers. On the one hand, there is a select set of researchers who take the lead: 'The head of department is responsible for ideas and research (results) and buys resources and support services from resource departments' (document 3). Reasoning from the perspective of the heads of department, a resource is defined as the 'capacity that facilitates or supports the execution of research ideas and projects or that independently engages in activities for the benefit of research' (document 3). On the other hand, there are resource departments who offer goods and services to researchers. Reasoning from the perspective of these service providers, a resource is defined as a 'service or activity for which there is an internal transfer price' (document 3). In building a rationale for algorithmic allocation, the exchange of resources was presented as a key organizational principle.

In addition to exchange, algorithmic allocation builds on the idea of 'achievement through competition' which ensures that the best performing researchers are rewarded while those who perform less well will see their resources decrease. As a matter of fairness, performing in a competitive environment is about doing 'justice' to what someone has accomplished: 'The idea of performance-based financing is an allocation of the benefits that does justice to your achievements' (financial controller 1). The value of competition is also visible in the decision to exclude resources attracted from pharmaceutical companies in the indicator 'acquisition power.' The Centre admitted that acquiring money from industry was an achievement in itself but argued that money from scientific and patient organizations still had a different status because it was 'acquired in (scientific) competition' (document 3). Competition is more than a feature of research management. The long history of algorithmic allocation, one professor says, tells us something about ourselves as researchers too in the sense that 'we are competitive but we do also want to be rewarded for our performance' (professor 4). The recent decision to restrict algorithmic allocation, for instance, thus led to feelings of wrongdoing: 'But now that it is reversed so much you see that we, as staff, say: "Well, yes, if your performance is not financed that is quite difficult too. Because that does not feel right"' (professor 4). In elaborating on that feeling, (s)he adds: 'Well, we too often had the feeling that in particular the people who would grow or did well would get into trouble. Plus…you want to have the feeling that you are rewarded for achieving something' (professor 4).

A third, motivational justification of market fairness is about 'offering incentives for innovation' to departments such that they will be induced to strive for excellence under similar conditions. The explicit objective of algorithmic allocation is 'to reward and stimulate good performances' and to offer 'an incentive for the optimal organization' of the Centre's research infrastructure (document 1; document 4). Under the heading of 'steering on incentives,' the medical Centre spoke of the need to steer on 'the quality of the output' and 'the societal impact of research' as well as the ability to stay 'within the research budget' and be 'in line with the focal areas of the centre' (document 3). Under that same heading, incentives were evoked as a reason to include citation scores of publications in the allocation model: 'As a firm, we wanted a higher MNCS. So how do you do that? Well, you will be steering your department in that direction' (financial controller 1). Again, incentives are not just a matter of research management. They are also evoked by researchers themselves when they express the need for an environment in which innovation is stimulated instead of hampered:

> You just have to strike the right balance because you do need the incentives such that people…that people are not part of a too big organizational unity that becomes too inert and bureaucratic. You do want to have the incentives to be innovative and to set up new things. (professor 4)

To be induced to innovate fairly requires uniformity in the opportunities to conduct research. In that regard, incentives also offered an important justification for charging for the overhead costs. This should happen uniformly throughout the organization because departments with more clinical duties might place their healthcare personnel in rooms officially reserved for research if these came at no cost. As one administrator recalls: 'Yes, we have tried beforehand to avoid these kinds of perverse incentives, as we called them at the time, not to strengthen these incentives too' (financial controller 1). All in all, incentives figure prominently in the attempt to build a rationale for algorithmic allocation.

There are clearly pragmatic reasons to opt for an algorithmic model to allocate resources for (bio)medical research. The financial reorganization in healthcare, for

instance, was an important pragmatic factor for charging for the overheads in research too. The weight of acquisition power in the model, moreover, is not just to induce researchers to attract as many external grants as possible—though that is certainly part of it—but also because these grants have to be matched to internal resources in order to be viable (document 3). But placing these pragmatic reasons aside, this section demonstrated how both administrators and researchers justify the need for algorithmic allocation in market-oriented terms: to break with the alleged unfairness of historical budgets; to maintain the idea of achievement through competition; to incentivize innovation under similar conditions, and to do so in an organizational context of exchange where prices and responsibilities are well-defined.

5. Challenging Algorithmic Allocation: Epistemic Fairness

Algorithmic allocation became controversial soon after its implementation. The biomedical departments, in particular, had difficulties in attracting external grants and their resources further declined—due to the weight of acquisition power in the model—until their very survival was at stake (financial controller 1 and 2; professor 2). In 2018, the new board of directors decided to reduce the variable budget from 50 to 10 percent and to limit departmental competition (document 2). The decision to put algorithmic allocation on hold—instead of letting departments go bankrupt—shows that market fairness only reaches so far into matters of research management. In fact, researchers and administrators have brought a range of issues to the fore to problematize algorithmic allocation and its normative underpinnings. These considerations of 'epistemic fairness' all hinge on the idea that there are differences in the practices of producing scientific knowledge (Borgman, 2017; Franssen et al., 2018; Hessels et al., 2019; Knorr-Cetina, 1999). Or as one professor expressed it: 'Yes, I do not mind the discussion about benchmarking, about comparisons. You have to be proud of your output. But you have to look a little bit at the diversity, at how science is done' (professor 1). Biomedical scientists, for one, work primarily in laboratories on a wide range of biological processes at the level of cells, proteins, and DNA. As preclinical or 'basic' research, biomedicine is often quite far removed from direct healthcare applications and requires a lot of equipment, animals, and chemicals. The health sciences, on the other hand, are a less capital-intensive branch of research—computers and databases suffice here. Engaged in the science of large numbers, health scientists study the association between health-related variables in large cohorts of people. The health sciences have a role in some screening programs but are often more aligned with prevention—and hence with regional and national policymaking—than with clinical practice. Such epistemic differences lead to four considerations of fairness that problematize—or directly challenge—the market-oriented justifications which have been analysed so far.

To start with, the various departments differ significantly in publication output and their typical citation impact. In certain fields, individual journal publications can be produced with relatively little effort. Writing an article can, for example, entail performing a set of relatively standardized statistical analyses applied to data that were not generated by the authors themselves. This contrasts with the significant amounts of effort that can go into conducting more practical experimental studies performed in other fields: 'If you are doing cell biology, it's extremely difficult to get a paper. And if you work in big databases, then you can get twenty papers. So, I think especially universities should understand and acknowledge that' (junior researcher 1). The administrative staff is also aware of differences in publication practices:

> As a joke I always say, I exaggerate of course: You have a fundamental-orientated department who work their butts off for one publication. And you have a life science department who can put its computer on in the morning and, hop, there we have another publication from the database….And what you notice then—that is, if you don't discriminate between different types of research—is that fundamental research is being eaten up because they have less output. (financial controller 2)

Our respondents also touched on more subtle aspects of epistemic fairness in the use of publication-related metrics. One administrator, for example, mentioned the example of a field whose relatively small-scale organizational structure has resulted in very high MNCS scores. Conversely, in some cases, citation scores are normalized for a broader set of specialities that can make the output of a given department look underwhelming in comparison.

> If you are in a niche-field like child medicine, to name one, you never have a high MNCS score because you are competing in internal medicine, you know. So they can never score on this….So in that sense, it is unfair competition. (financial controller 2)

Furthermore, some of our respondents were concerned with the fact that a focus on output and citation results makes cooperation among departments less visible: 'But you can ask yourself: Okay, this department does not publish that much but they are actually pretty important for the medical centre because they have this specific expertise that many other departments, in fact, do make use of' (professor 2). A department, for instance, might provide data or other resources that are crucial for the work of others but 'if you are stuck to performance-based financing [then] you are at a point of losing such a department' (professor 2).

Secondly, scientists also clearly experience the competition for a larger market share as something that exerts a certain pressure on them: 'The department gets a financial kind of award, I guess, per publication, right?…So that's where the pressure starts, right?' (junior researcher 2). In their eyes, this leads them away sometimes from fair and responsible behaviour. For one, it informs publication practices and increases the rivalry over authorship positions:

> There is the performance-based financing, so it will get worse and worse because we get the [name scientometric institute], the citation, the MNCS. For our department, for example, it will be much better to publish in lower impact [journals], because it is too difficult [otherwise], and have more papers. And then we should claim all the first and lasts because otherwise our department will lose money for research and we have to fire people. (senior researcher 1)

In the focus group setting, another senior researcher responded by saying that the allocation model leads to a very particular kind of creativity: 'We all are creative then to make sure we get this performance-based financing, this additional funding. But is it really responsible science? That's the question' (senior researcher 2). One of the key ways to be responsible is to address a (scientific) audience that your results are most relevant for. Due to the exclusive focus on publications in the highest quartile of journals, however, other kinds of considerations easily take over. One of the things that (bio)medical scientists reflect upon is the instrumental way in which they seize upon epistemic differences between fields: 'Q1 is determined for a disciplinary field. So you shop around there. So sometimes it is: Oh, now I need to publish in a biology journal because there I can still publish in Q1, whereas in immunology it is not possible' (junior researcher 3). Adding to this practice of 'shopping around,' another researcher points to a tension between publishing in higher-ranked journals and finding a suitable audience:

> But then, if you think back on the responsible researcher and you shop around because you want to publish in the Q1 journals. But it might not be the journal which is most suitable for your research and with the readers….It is something strange you do.' (junior researcher 4)

Differences in the perception of the value of different publication outlets—again bound up with field differences—also permeate the relationship between junior and senior researchers:

> For me, it's actually quite tricky….For my future, it is better to publish in more scientific journals. But it's easier to publish in Q1 in less scientific journals. So my PI [principal investigator] wants to publish in these journals while I want to publish in the more scientific journals. (junior researcher 1)

In reflecting upon daily decision-making on the scientific shop floor, the algorithmic allocation model is explicitly related to the struggle over authorship positions and to attempts to 'creatively' seize upon epistemic differences in publication and citation patterns between different (bio)medical fields. The placing of authorship within the allocation model at least gives institutional legitimacy to the (perceived) pressure to secure good authorship positions in order to add to the market share of the department—but it might equally contribute to a struggle over authorship positions.

A third important aspect of epistemic fairness relates to an unequal ability of departments to attract resources beyond the confines of the medical Centre. The reason for this is the highly diverse structure of academic and commercial funding opportunities across fields.

> For [some] discipline[s] it is way easier to acquire external funding. Of course, there is a lot of competition but the total availability of funding for oncology, for example, is much larger than for medical ethics or plastic surgery. How can you use the same measuring rod? (research policy advisor 1)

One interviewee contends that differences in the access to external resources make for 'a really difficult balance' in the allocation model because 'some research can get money more easily than other research' (professor 4). This interviewee perceives this as having to do with how 'sexy' the research is: 'I always say that you just have an easier time when you do something with pitiful children or cancer' (professor 4). More fundamentally, access differs between clinical research on the one hand and basic research in biomedicine and the health sciences on the other. Research with greater relevance for clinical practice, for instance, is depicted as a 'completely different branch of sports' due to the access that the clinical departments have to the resources of pharmaceutical companies:

> [Y]ou also have to realize…that big pharma is on top of it; it pours a lot of money into it. And a lot of money goes specifically to these departments. Yeah, they can, of course, deliver these pharma outcomes, but they can also partly use that money and fund other research projects that would maybe not get funded otherwise. (professor 2)

A further complicating factor is that according to some of our subjects, more fundamental forms of research tend to encounter difficulty in convincing patient organizations of the value of their work. This was perceived as an additional source of imbalance in the overall distribution of funds across (bio)medical research, with the current algorithmic allocation system reinforcing the problem:

> And in my field—I'm working on a pretty molecular level, so patients, of course, don't see the relevance in that….So what's happening now, looking at all the VENI's and VIDI's [major Dutch research grants], is that a lot of proposals closer to the clinic get the money because, whatever….And now we don't really notice the effect but in ten years we will definitely notice it. (junior researcher 1)

Fourth and finally, our interviewees point to differences in the average duration of doctorates between fields and terms of how successful departments can claim to be in supervising graduate students. As a general rule, doctorates in clinical fields tend to take less time than in biomedical fields:

> Concerning doctorates, that takes five years on average at a biomedical department while it takes three years in the clinic….So you actually have very specific norms per group or per research field about what you need to really do. Because if you have the same measuring rod for everything, then, yeah, you destroy more than you achieve according to me. (financial controller 1)

Interviewees account for such differences by referring to the greater degree of unpredictability of laboratory-based research. One professor explained that unexpected outcomes of experiments are common and can significantly delay graduation or even lead students to give up while this is less common in the more clinical areas:

> Look, we, of course, have departments…where it is way easier to channel PhD students. It is a kind of doctorate factory. And in the laboratory, it is plodding on and if something fails then a whole doctorate might fail. That is quite a difference. (professor 4)

In summary, the considerations presented in this section draw on perceived differences in the practices of knowledge production to confront the quantitative indicators at the heart of algorithmic allocation (see also Rushforth & de Rijcke, 2015). Following this line of reasoning, departments no longer appear as organizational units that can compete fairly based on a set of uniform performance indicators covering doctorates, external grants, and the number and scientific impact of publications.

6. Conclusions

In our article, we addressed an analytical question central to higher education as it is managed and experienced nowadays: What kinds of normative consideration do researchers and administrators draw upon to justify or challenge algorithm-based allocation? To answer that question, we untangled two rival notions of fairness that were at play in research management. The long history and continuous importance of performance-based financing at our university medical centre first led us to four considerations of market fairness: Eliminating the arbitrary interference that has previously privileged some departments at the cost of others; installing achievement through competition as insurance that the best departments were rewarded while those who performed less well saw their resources decrease; introducing incentives to induce innovative research and to stimulate departments to excel under similar conditions; and reconfiguring the research organization as a site of exchange where people pay for the services used and receive money for the things they deliver.

The decision to put algorithmic allocation on hold led us to the second series of normative considerations that problematized the idea that it resulted in a fair distribution of resources. These considerations of epistemic fairness, as we called them, build on the idea that a university medical centre might be unified in its focus on health and disease but very diversified in its cultures of knowledge production and their affordances and constraints (Borgman, 2017; Franssen et al., 2018; Hessels et al., 2019; Knorr-Cetina, 1999). Differences in the publication practices between different (bio)medical fields was the first normative issue brought to the fore as a challenge to the alleged fairness of the allocation model. In addition, researchers linked the insistence on Q1 publications and the importance of authorship positions in the model to struggles between researchers to secure the best position and to tensions at the heart of their decision to publish in one journal or another. Third, researchers and administrators pointed to differences in the ability of departments in different fields to attract external resources. Finally, they singled out the 'doctorate' as a research trajectory that had a different temporal and risk-related meaning for different PhD students and departments. When we bring these considerations together, the organization appears as something more than a site of competition, incentivization, and exchange. Instead, those with decision-making power should account for differences in publication practices, doctoral research, and access to resources, and acknowledge that research is not just a matter of competitive achievements of otherwise homogeneous departmental units.

The case of algorithmic allocation at a university medical centre does not stand on its own. It fits in with broader trends of quantification, marketization, and algorithm-based decision making in higher education—as well as in society more broadly. On the one hand, our analysis corroborates earlier findings about the importance of indicators and rankings for research management as well as the spread of market-oriented ideals and practices in the current organization of research (Hammarfelt et al., 2017; Paradeise & Filliatreau, 2016; Rushforth & de Rijcke, 2015, 2017; Sauder & Espeland, 2009). By untangling considerations of market fairness, we were able to better understand the appeal of competition and quantitative performance indicators for allocation purposes. On the other hand, however, our analysis of epistemic fairness shows that we should be

careful in depicting higher education as populated by 'market universities' fully enmeshed in 'epistemic capitalism' (Berman, 2012; Birch, 2017; Fochler, 2016; Slaughter & Rhoades, 2009). Although these trends definitely influence research organizations, they remain heterogeneous institutions in which rival normative considerations are at play (Paradeise & Thoenig, 2018; Rushforth et al., 2018; Whitley, Gläser, & Engwall, 2010). Beyond the specifics of the case, our analytic distinction between market fairness and epistemic fairness thus provides insights into the relative instability and (potential) room for manoeuvre of research organizations. The normative tension we discerned offers a way into the internal dynamics of organizational debates and—possibly—a way out of the limitations that marketization, quantification, and algorithms impose.

The current state of the allocation model at the Centre provides an interesting test case of how to deal with normative tensions in a broader 'algorithmic culture' (Striphas, 2015). The future of performance-based financing at the centre is still undecided. Algorithmic allocation is currently severely restricted, but the board of directors has tasked a special committee to reconsider performance-based financing and there is a push from the heads of department to slowly bring it up to its former level. The broader literature on computation and algorithms could offer interesting perspectives for consideration by decision-makers in higher education and elsewhere (Berman & Hirschman, 2018; Finn, 2017; van Dijck, 2013). We flag two specific sources of concern. The first concerns the way institutions should sustain wider legitimacy and be held 'algorithmically accountable' (Couldry, 2016). The partial or full delegation of human decision making to automation raises profound questions about responsibility and accountability for the outcomes of these decisions. Are decision-makers fully able to understand the algorithms and their consequences? And can they still explain to others that the decision made sense when they are prompted to do so? The second issue concerns the way to deal with rival considerations of fairness in designing, implementing, and reviewing algorithms. Our case demonstrates that there are crucial differences in the way people justify or problematize algorithm-based decision making. Can we expect algorithms to ever resolve the normative tensions that divide organizations and human societies? Or do algorithms aggravate such tensions because some considerations are more easily embodied while others are pushed to the side? These two sources of concern, accountability and normativity, are not so easily addressed. But we do expect them to appear and reappear whenever and wherever algorithm-based decision making predominates—in higher education institutions as well as in other areas of social life.

Acknowledgments

The research conducted for this article was funded by The Netherlands Organisation for Health Research and Development (ZonMW). We would like to thank the reviewers for their critical and constructive comments.

Conflict of Interests

The authors declare no conflict of interests.

References

Berman, E. P. (2012). *Creating the market university: How academic science became an economic engine*. Princeton, NJ: Princeton University Press.

Berman, E. P., & Hirschman, D. (2018). The sociology of quantification: Where are we now? *Contemporary Sociology*, *47*(3), 257–266.

Birch, K. (2017). Rethinking value in the bio-economy: Finance, assetization and the management of value. *Science, Technology & Human Values*, *42*(3), 1–31.

Boltanski, L., & Thevenot, L. (2006). *On justification*: *Economies of worth*. Princeton, NJ: Princeton University Press.

Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.

Couldry, N. (2016). What's at stake in algorithmic accountability. *Media@LSE blog*. Retrieved from https://blogs.lse.ac.uk/medialse/2016/04/28/whats-at-stake-in-algorithmic-accountability

Finn, E. (2017). *What algorithms want: Imagination in the age of computing*. Cambridge, MA: MIT Press.

Fochler, M. (2016). Variants of epistemic capitalism: Knowledge production and the accumulation of worth in commercial biotechnology and the academic life sciences. *Science, Technology & Human Values*, *41*(5), 922–948.

Fourcade, M., & Healy, K. (2007). Moral views of market society. *Annual Review of Sociology*, *33*, 285–311.

Franssen, T., Scholten, W., Hessels, L. K., & de Rijcke, S. (2018). The drawbacks of project funding for epistemic innovation: Comparing institutional affordances and constraints of different types of research funding. *Minerva*, *56*(1), 11–33.

Hammarfelt, B., de Rijcke, S., & Wouters, P. F. (2017). From eminent men to excellent universities: University rankings as calculative devices. *Minerva*, *55*(4), 391–411.

Hessels, L. K., Franssen, T., Scholten, W., & de Rijcke, S. (2019). Variation in valuation: How research groups accumulate credibility in four epistemic cultures. *Minerva*, *57*(2), 127–149.

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Paradeise, C., & Filliatreau, G. (2016). The emergent action field of metrics: From rankings to altmetrics. In E. Berman & C. Paradeise (Eds.), *University under Pressure: Research in Sociology of organizations series* (pp. 87–128). Bingley: Emerald Group.

Paradeise, C., & Thoenig, J. C. (2018). Higher education institutions as strategic actors. *European Review*, *26*(Supp. 1), 57–69. https://doi.org/10.1017/S1062798717000540

Rushforth, A. D., & de Rijcke, S. (2015). Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands. *Minerva*, *53*(2), 117–139.

Rushforth, A. D., & de Rijcke, S. (2017). Quality monitoring in transition: The emerging challenge of evaluating translational research programs in academic biomedicine. *Science and Public Policy*, *44*(4), 513–523.

Rushforth, A. D., Franssen, T. P., & de Rijcke, S. (2018). Portfolios of worth: Capitalizing on basic and clinical problems in biomedical research groups. *Science, Technology, & Human Values*, *44*(2), 209–236.

Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review*, *74*(1), 63–82.

Slaughter, S., & Rhoades, G. (2009). *Academic capitalism and the new economy: Markets, state and higher education*. Baltimore, MD: Johns Hopkins University Press.

Striphas, T. (2015). Algorithmic culture. *European Journal of Cultural Studies*, *18*(4-/5), 395–412.

van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford: Oxford University Press.

Whitley, R., Gläser, J., & Engwall, L. (Eds.). (2010). *Reconfiguring knowledge production: Changing authority relationships in the sciences and their consequences for intellectual innovation*. Oxford: Oxford University Press.

About the Authors

Guus Dix is a Postdoctoral Researcher at the Centre for Science and Technology Studies. He works on a ZonMw-funded project on fostering responsible biomedical research and an ERC-funded project on science policy and ocean science. In the historical and sociological book project *Incentives: A Genealogy* (under contract with Princeton University Press), he traces the development of 'incentivization' as a framework to understand human behavior and set of instruments to steer it in a different direction.

Wolfgang Kaltenbrunner is a Researcher at the Centre for Science and Technology Studies at Leiden University. Wolfgang's current research focus is on the politics of quantification in evaluation processes, on academic curricula vitae as devices for comparison in peer review, and on practices of publishing, reviewing, and editing academic literature. Wolfgang's research has appeared in journals such as *Social Studies of Science*, *Minerva*, and *Science & Public Policy*.

Joeri Tijdink is a Senior Researcher and Psychiatrist affiliated with the Vrije Universiteit Amsterdam and AmsterdamUMC. He undertook several research projects to uncover the major causes of research misbehavior. His special interest is in the role of publication pressure and personality traits as a risk factor for research misbehavior. Currently, he is involved in several research projects that investigate research misbehaviors, research culture, the importance of mentoring for early career researchers. With an international group, he creates institutional guidelines to foster responsible research conduct.

Govert Valkenburg is a Senior Researcher in Science, Technology and Society Studies at the Department of Interdisciplinary Studies of Culture of the Norwegian University of Science and Technology. He currently studies knowledge dynamics in innovation processes and connects them to political theories. His research has covered energy transitions, privacy and security technologies, and human genomics. He was recently awarded a personal Researcher Project grant by the Research Council of Norway, for research into knowledge equity in the securitization of critical infrastructures.

Sarah de Rijcke is Professor in Science and Evaluation Studies and Scientific Director at the Centre for Science and Technology Studies in Leiden. Chair of the research group. She also is Co-Chair of Research in Research Institute (RoRI). Sarah specialises in social studies of research evaluation, and has published widely on the topic of the relations between quality control mechanisms and knowledge production in different fields. She has also developed a strong international public academic presence.

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2020, Volume 8, Issue 2, Pages 26–35
DOI: 10.17645/pag.v8i2.2582

Article

Steered by Numbers: How Quantification Differentiates the Reform of a German University

Michael Huber

Department of Sociology, Bielefeld University, 33615 Bielefeld, Germany; E-Mail: michael.huber@uni-bielefeld.de

Submitted: 29 October 2019 | Accepted: 31 December 2019 | Published: 9 April 2020

Abstract
Quantification theories assume that numbers govern and steer a policy field or an organisation. In order to steer successfully, however, the local interpretation of numbers takes centre stage as the meaning of numbers—and thus the way how actors respond to them—varies between systems or sectors. Empirically, this article reviews how a German university makes sense of political numbers and their implicit steering signals, and how quantification alters its organisational structures and reshapes the roles of academics. The article analyses the translation process distinguishing between three levels: the political discourse on university reform; the organisational adaptations; and the effects they have on the professional academic role. The article finds that the university has highly differentiated strategies to respond to the 'governance by numbers,' and that it has established independent number-based steering systems. We also find that such differentiation of programmes makes the university management more flexible, helping it deal with anticipated goal conflicts and unwanted allocative effects, but it also places serious strain on—and potentially overburns—the coordination provided by the university's central administration. We also find that academics have started to align their behavioural strategies towards fulfilling their organisational goals and that they tend to deviate from professional expectations. Discussing these differentiated strategies, this article shows how the differentiation of governance approaches also contributes to the university becoming an 'organisational actor.' These preliminary findings suggest the need for and potential direction of further investigations.

Keywords
organizational reform; quantification; university management; university studies

Issue
This article is part of the issue "Quantifying Higher Education: Governing Universities and Academics by Numbers" edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

1. Introduction

For the past 40 years, higher education studies have become acutely aware of and have criticized, the various changes and reforms in the sector. In this context, the new public management (NPM) provided the main analytical lens through which reform was perceived and practical tools to overcome the diagnosed inefficiencies, opacity, and professional idiosyncrasies of the higher education sector were chosen (for NPM see Hood, 1991). Over the years, conceptual and practical innovations have been added, suggesting that reforms do not form part of a grand plan, but recurrently intervene to repair the shortcomings and unintended effects of previous interventions (on this general feature of reform see Luhmann, 1992, p. 74). Wittrock summarises this *layering* nicely: "Universities exist with layer upon layer of quite divergent legacies, yet somehow they have also succeeded in preserving a strong element of continuity amidst all the change" (Wittrock, 1993, p. 305). The most current layer of university reform is 'quantification,' which means that numbers are behind identification and adoption of promising strategies, permit the 'objective' observation of the effects of reforms, and ensure accountability. This 'governance by numbers' (e.g., Miller, 2001; Rose, 1991) is expected to rationalise and modernise the university, and—for example through mechanisms of rankings (see Ringel, Brankovic, & Werron,

2020)—the entire higher education sector. However, the main assumption of this article that this rationalisation process does not come without costs; it not only restructures the university but also reshapes the university's relation to its environment, i.e., politics and staff members.

In the theoretical narrative of this new layer, the effectiveness of numbers is taken for granted while the documented experience with the numerical steering of organisations, and more specifically, universities, is at best ambivalent. One widely communicated reason refers to the resistance of the university to any change; in particular, the German university has a longstanding reputation of being unable and unwilling to reform (e.g., Stichweh, 1994). Another, methodological reason, which the wider public are often concerned with, relates to the accurate measurement of professional performance. This article explores how governance by numbers actually occurs, if or to what extent universities live up to the implicit promise of quantification's effectiveness, and, most importantly, this article asks how numbers restructure the organisation.

The argument is outlined in three sections. First, the quantification literature is reviewed in order to show how numbers can improve steering. As far as 'governance by numbers' is concerned, Miller and O'Leary's (1994a, 1994b) account of the reform at Caterpillar in the 1980s is used as an analytical frame for the empirical exploration of the numerical reform of universities. Examining the effects of numbers on organizational reform, the authors suggested to not focus on the organisation alone, but to include the societal or political discourse as well as to reflect on the alterations at a personal level. Thus, Miller and O'Leary's blueprint allows us to describe university reform by discerning three intertwined analytical levels: First, the political programme promoting generic ideas about the need, and main direction, of reform. Second, the organisational level, where these programmes are enforced by changes to the production process and the reallocation of responsibilities. At a third, individual level, the workers are assigned new responsibilities and skills for the production process. Numbers imply that workers need to develop numeracy, i.e., the skill to understand and apply numbers. This layered model has been created for a single-case study and used to develop a broader understanding of reforming economic firms through numbers. This article explores the process of quantifying the performance of universities and compares how quantification reshapes the 'production processes' of a German university. This exploratory study leads to a set of tentative questions that should help to design further systematic research.

2. Quantification and Organisation

This section is divided into three subsections: First, we briefly sketch a communicative perspective on quantification; second, we outline the basic argument of the Caterpillar study by Miller and O'Leary. Together these considerations allow us to, third, delineate a framework that structures the empirical investigation of Section 3.

2.1. Quantification and Organisational Change

Even at the turn of the last century, Weber (1978) already considered accounting and control through numbers to be a decisive step in the development towards the modern capitalist society. The broader social phenomena we subsume under the notion of quantification, however, have long been ignored by the social sciences (overview in Miller, 2007). This neglect was turned into awareness only over the last two decades when numbers were identified as critical for modernity (prominently, Porter, 1996) and the establishment of state organisations (e.g., Heintz, 2012). Increasingly, numbers are identified as a ubiquitous phenomenon (Mennicken & Espeland, 2019). As a result of this growing interest, quantification studies provide a wide, internally highly differentiated and specialised research area (overview in Heintz, 2018). This converges in the broad definition that quantification comprises "the production and communication of numbers—and its consequences" (Espeland & Stevens, 2008, p. 402). Such a communicative approach to quantification is anticipated by systems' theory, as Luhmann (1997) had already suggested before the excitement over quantification, that numbers simplify communication as they insulate the respective statement against criticism; in other words, arguments sustained by numbers are more difficult to negate than others (see also Heintz, 2016). This basic idea also serves as a starting point for the debate on quantification by Espeland and Stevens (2008). These authors, however, further develop the basic communicative model of quantification with the help of the theory of speech-acts by John Austin (1975), and discern different types of numbers and how numbers may improve and rationalise governance. Assuming that different things are done by different numbers (Espeland & Stevens, 2008, p. 405), the main focus of Espeland and Stevens is on the varied consequences of quantification, i.e., how new categories of countable things emerge, how numbers and their implicit accuracy can be interpreted, and where and when new infrastructures of counting materialise. One aspect of Espeland and Stevens's diversity-of-numbers argument that's of particular interest to this article concerns the contextualised meaning of numbers. For example, when universities use publication scores to document scholarly productivity (see Krüger, 2020), these scores assume a different meaning when they are applied, e.g., by funders to benchmark resource allocation. The effects of this differentiation of meanings are also investigated by Power (2007). He proposes a 'sequential hypothesis' about the changing meaning of numbers: Numbers are first generated at the operational level to make performance tangible and visible for management. These numbers 'turn-the-inside-out,' providing evidence of the organisational

performance from the perspective of operators. Once these numbers are released into the public realm, they can be used at will, for example, to promote external—i.e., mainly political or economic—interests, and then be redirected to the operational level. This 'turning-the-outside-in' may challenge the local (often professional) performance routines. These effects of quantification on higher education have, for example, been studied in Espeland and Sauder's (2007) seminal article on reactivity. Their study suggests that the university management responds to results of rankings by improving their performance but also by gaming those numbers in order to climb up the ranks. As far as the active steering of organisations through numbers is concerned, the practice should, generally speaking, rationalise management and unburden the bureaucratic authority as "members are provided with information, which, when previously programmed benchmarks are met, triggers decisions" (Luhmann, 1964, p. 98, author's translation). Empirical studies on the effects of numbers on universities suggest this automated steering is not without collateral effects (e.g., Huber & Hillebrandt, 2019).

The basic argument is that numbers call for interpretation (or translation) at the point of use, otherwise their unfiltered application may be incomprehensible to the addressees. Thus, the general promise of a communication-based concept of quantification is to efficiently implement reform strategies, albeit at the cost of altering the organisation, its perception of external (and internal) demands, as well as its decision-making procedures. This article explores the specific forms and empirical variations in performance indicator communication within a German university—Middletown University (MU). This article uses Miller and O'Leary's (1994a, 1994b) analysis of the reform of the Caterpillar plant in Decatur, Illinois, as its analytical framework which is discussed in the next section.

2.2. The Caterpillar Story: An Analytical Framework

In the early 1980s, the Caterpillar company, which designs, develops, and produces machinery and engines, failed to compete successfully with its Japanese competitor Komatsu. Miller and O'Leary (1994a, 1994b) analysed Caterpillar's recovery and showed that the process of making the company profitable again started at the political level. The political discourse perceived Caterpillar and its problems as part of a broader challenge to the American industrial sector, its competitiveness, and more generally to American prosperity itself. The modernisation and reform of US industries in general, and of Caterpillar in particular, called for plant re-design and re-training of the workforce; its main objective was, from a political perspective, the re-establishment of American competitiveness. Although not all voices in the political discourse agreed on every aspect of how the American industry should recover, they all converged on the 'need to act.'

Against this political background, the Caterpillar company re-structured and modernised its plants. Miller and O'Leary focus their studies on the plant in Decatur, Illinois, where two dimensions of organizational adaptation caught their attention: First, the production process was re-organised through a new assembly line that simplified and modularised the production process—the authors also refer to "cellular working arrangements" (Miller & O'Leary, 1994b, p. 480). The second dimension concerns the establishment of a quantitative 'audit trail' for all aspects of the production process enabling continuous comparison with the Japanese competitor through numbers; this "competitor benchmarking" is a:

> Calculative practice that the image of Japanese competition was made real to those working in the North American plants of Caterpillar Inc. By this means, 'competitiveness' was no longer an abstract idea, a simple invocation to work harder, to do more, to produce quicker. *'Competitiveness' meant 'person-to-person' competition with a Japanese worker.* The "threat" from Japan to American manufacturing was to be given a face, *and* a number. (Miller & O'Leary, 1994b, p. 472, author's emphasis)

This practice enabled the comparison of all activities within the plant, between plants, and their competitors. This comprehensive comparability also changed the expectations of the skills of the workforce. Benchmarking forced Caterpillar to re-educate its workforce as it—together with the modularised production structure—established "a new way of relating individuals to their work within the factory" (Miller & O'Leary, 1994b, p. 477). Workers had to develop numeracy, i.e., the continuous comparison with the competitor's performance required them to learn how to read numbers and then to have the flexibly to adapt their work to diverse benchmarks. At this individual level, quantification established an 'economic citizenship' that not only empowered workers but also placed new responsibility on their shoulders, and thus altered their traditional membership role (Miller & O'Leary, 1994b, p. 478). Miller and O'Leary (1994b, p. 473) notice that the workforce obtains a new, more influential role, as:

> Authority would flow directly from the customer to the work process, along the Assembly Highway, in accordance with the ideal of empowered workers responding immediately to the wants and wishes of the customer. Authority would no longer be embodied in the character of the supervisor, or in the routine calculations of a technique such as standard costing, but would inhere in the process itself.

As mentioned above, Luhmann (1964, p. 98) suggested that quantification unburdens the management. Miller and O'Leary confirm this assumption and add that the burden is internally re-arranged and, in large parts,

shifted to the workforce. These changes, however, do not establish a stable structure. Instead, Miller and O'Leary (1994b) emphasise the *instability* of the new configuration of the production process:

> A perpetually failing series of programs and instruments for governing economic life. It is this *instability*, combined with a constant search for temporary stabilities, that in large part explains the process of trying to create a new reality on the factory floor in a particular North American factory in the 1980s. (p. 491, author's italics)

Instability, and, in response to it, the continual repair of failures, is seen as a critical part of the success of Caterpillar's reform as it established a permanent adaptation process. The organisation, staff, and politics were kept alert at all times.

2.3. Guidelines of Empirical Exploration

Miller and O'Leary's (1994a, 1994b) layered model of organisational change provides us with a basic understanding of reform that suggests, first, a distinction between three levels, i.e., politics, organisation, and individuals. Reflecting on the interactions of these three levels is expected to more comprehensively capture the drivers of, and challenges to, reform. Second, the reorganisation of the working procedures draws particular attention to the receiving side of knock-on effects, i.e., the workers and their skills to handle quantification and to assume new responsibilities are considered important although they are often overlooked in the literature. Third, the instability of reform is considered essential for its success, i.e., success is based on establishing what we described as a continual process of repairing the flaws of the previous reform (see also Luhmann, 1992). These basic elements of an analytical framework guide the empirical analysis of the next section.

3. The Case of a German University and Its Reform

This study follows the single-case-scheme outlined above. Thus, the results of the empirical part are tentative and exploratory, focusing on the effects of quantification at the successful, medium-sized MU in Germany. The MU lacks a medical department and instead has developed a strong social science focus. Such deviation from the 'normal university' requires more managerial effort by the MU-management as the state-programmes discussed below are designed to steer a comprehensive, 'normal' institution. The study is based on policy documents and interviews. References to MU-documents and interviews have been anonymised and the cited passages from the interviews were translated by the author. At the core of the article, an interview with the team of the Vice Chancellor's office, including the Vice Chancellor, is analysed sequentially (e.g., Herz, Peters, & Truschkat, 2015); it explores what specific solutions to the problem of quantification and reform can be identified at MU. The focus on one university clearly suggests that the main results of this article cannot be 'representative.' At best, it is able to identify critical issues which could inform and further guide more systematic investigations into the relationship between quantification and organisational reform across German universities.

3.1. The Political Side of the University Reform

In the early 1990s, the political debate surrounding German higher education emphasised the urgency of university reform. The reform needed to respond to three principal challenges. First, the need for the regime to be more efficient in order to provide reasonably priced higher education for as many people as possible. Second, the once internationally leading German university system had become a laggard in global competition, failing to innovate and thus endangering the prosperity of German society. A third challenge concerned the academic profession. The professional opacity turned decision making into a 'garbage can' (Cohen, March, & Olsen, 1972), it protected "rotten apples" within the academic profession and avoided to hold professionals accountable for the substantial resources used in the higher education sector (for the German case see e.g., Schimank, 2005). Political discourse at the time highlighted the need for change; the NPM-doctrines were expected to provide the tools necessary to improve the efficiency and accountability of universities and consequently reshape the welfare and competitive functions of the higher education sector.

However, the German reform discourse takes us a step further as it converges at the assessment that the university is unable (and often unwilling) to reform itself. Thus, reform has to be driven externally, by the state (Stichweh, 1994). The traditional state-dependency of the German universities (see also Clark, 1986) could be identified in the interviews. The interviewees addressed the state and its reform objective right at the beginning of the conversation and summarised the basic reform idea: "Achievements should be rewarded, that was the key motto of the Lander Parliament" (interview).

This ambition to reward achievement suggests the introduction of differentiation in resource allocation to benefit those who perform better. The way this differentiation is implemented requires more concrete steps of operationalisation. One important step was shown at the beginning of one of the interviews, when the interviewers briefly introduced the comparative research project. One interviewee reacted to this methodological remark by reporting that MU was placed in a setting where all universities of the land were being compared according to performance indicators (see in next paragraph *Leistungsorientierte Mittelverteilung*, in short LOM). The state administration had delineated a comparative environment in which the reform unfolded. The state de-

fined the objects of comparison (universities) and the criteria of comparison (indicators), and together they indicate where and how much differentiation is in political demand (for these critical elements of comparison, see Heintz, 2016).

The state administration designed the LOM (i.e., 'performance-based budgeting'; for more detail, see Huber & Hillebrandt, 2019) which condenses the state's ideas regarding how and where the universities should compete. The LOM operationalises academic performance through publication scores, third party funding, and 'status'-indicators. While all of the academic activities are measured, the interviews emphasised that only a small part of the state's monetary allocations is used for such competitive arrangements. Within the cameralistic budgeting, LOM is therefore biased and relative. Biased, as LOM weighs performance uniformly across the sector and fails to account for differences, for instance, among the size of universities or disciplinary characteristics. For example, MU can be expected to have a weaker performance in publications and third-party funding as it lacks a medical department that normally performs well on these indicators. The university may counteract this bias by adapting and optimising the measurement methods, but a bias towards natural and life sciences, as well as bigger units, affects financial distributions. LOM is relative as resource allocation is restricted to single accounts of the cameralistic model, not the overall budget. This implies that the resource pool does not increase with increased productivity, but (if at all) as the result of political decisions. Therefore, redistributions by LOM do not reflect improvements in productivity, but the relative increase vis-à-vis other universities. In effect, one could get less money for higher productivity if other universities had increased their productivity to a greater degree. As a result, the LOM challenges planning at the organisational level (a repeated complaint throughout the interviews) as it does not address the immediate performers but affects the university as a whole. Another effect of this comparative setting is that the university has become the key object of comparison. In the higher education studies, this shift has been captured by the university "becoming an organisational actor" (e.g., Krücken & Meier, 2006). Practically, this means that the university administration is held responsible for the performance of its staff and the optimal positioning of the university in the LOM scheme. Control of performance is shifted towards the university. Besides LOM, the state also installed other programmes where additional resources are allocated. The interviewees mentioned that the state compensates for: (i) tuition fees; (ii) negotiates performances directly with the university management (for example, student uptake); (iii) adds overhead resources to third-party funding; and (iv) allocates resources directly to chairs or faculties. All these programmes are initiated by politics (see also Hillebrandt, 2020). They are fragmented and vary greatly in terms of objectives and allocative procedures, but their common feature is the idea of rewarding achievement through differential allocation.

The state imposes governance by numbers and the university steers some of these programmes with the help of their own numbers. For example, tuition fees were highly contested in Germany, and after 2013 all Lander administrations retreated from this idea and provided free higher education again. Given that fees transformed the income structure of universities, the state substituted the fees, however not as part of the overall budget, but as an isolated, publicly visible provision to accentuate students as an income source of universities. These direct allocations are based on an undifferentiated headcount but are turned into a performance-driven regime inside the university. Another state-programme directly allocates funding based on agreements with institutes or persons as direct service providers for the state; here experiments regarding student access issues, specific educational programmes, or expertise are initiated and supported through additional financial allocations. These allocations are partially driven by indicators but depend mainly on the task at hand. To some extent, they can be connected to quantification, but this connection is mainly made by the selection of providers. The next section outlines how the university 'internalises' these state-programmes.

3.2. How the MU Adapts the Political Reform

The state plays a critical role in university reform and uses quantification to operationalise some individual programmes. Major parts of the implementation of these programmes, however, are delegated to the university. This division of labour is recognised by MU and seen as being a part of the reform process: "Actually, prospective management is all the Land is providing, *it is up to us to provide the rest. And universities handle it differently*" (interview, author's emphasis).

Thus, the university's actorhood manifests itself in MU's new responsibility for performance and its translation of these 'state numbers' into locally meaningful and applicable strategies and indicator systems. The remainder of this section illustrates some examples of how this has been done at MU.

The Caterpillar model suggested that work processes need to be modularised, simplified, and restructured. But to internalise competition, Caterpillar not only restructured the process but also made its elements easy for the workers to compare through 'competitor benchmarking.' The MU case suggests a different development as the main source of change is not the organisation but the state-LOM and its predefined competitive environment. The MU mirrored and slightly adapted the LOM by making subtle changes. First, the university scheme changed the relative weight of LOM-indicators. Second, although most indicators are standardised across the sector—e.g., number of publications, student/teacher ratios, graduates, and third-party funding—local and po-

litical indicators such as 'gender equality' were added. Third, algorithms like the averaging the outcome over three years or installing caps to the possible changes of resource allocations were applied to ensure a robust, manageable outcome. The university scheme deviated from the LOM and the interviews highlighted that the university's foremost interest was not competition, but to guarantee a *manageable* allocation of resources. As far as the production process is concerned, the performance indicators address the Chair, i.e., the professor and her (academic and non-academic) collaborators, as the main production unit. At first sight, this arrangement seems unspectacular, but it gets more interesting when compared to other allocation strategies.

Student fees provide a different example of organisational adaptation. As mentioned above, all Lander parliaments ceased to demand fees from students after 2013. To substitute these payments, the state compensated a certain amount per student, i.e., it focused on the receiver of performance, and treated all students equally. At the university level, the management of student fees was translated into a very different idea that altered the production process and, in the course of the reform, established a new 'performance address.' The organisational strategy also started with the number of students but allocated their fees per 'full-time student equivalent' (*Vollzeitstudienäquivalent*, VSÄ), i.e., an indicator reflecting the relative share that teachers and teaching groups have in the education of each single student. Thus, the university shifted the focus onto the performance of teachers. The VSÄ reflects that students are not taught by one, but by several teachers, all of whom should get a fair share of the fee. Thus, the VSÄ places performance at the (aggregated) level of the module (and exams), as this is considered the *fairest* representation of teaching performance. Manageability is now considered to depend on internal fairness, otherwise 'system maximisers' could exploit the system and generate additional managerial challenges (for more, see Section 3.3). Thus, MU-teachers connected by the modules received their 'payment' relative to their share of each module; internally, all teachers linked to a specific module got an equal share of the fee. From an organisational perspective, membership of such professional groups became critical. An interviewee remarked on the previous situation:

> We had no legally responsible units in the departments. Professional groups were informal, nobody could determine who really belongs to them. Because, if somebody would say, I have got nothing to do with your professional group, I set up my own, nobody could have objected. (interview)

With the fees, the professional groups became the new addressees. They were easily rearranged and at least partially able to overcome the problems of internal redistribution in departments which were described as follows: "We had the experience that individual units, teaching units, allocate very different burdens….Single staff members supervised 30–40 diplomas, and one staff supervised one per term and refused to accept a second supervision" (interview). Thus, the use of VSÄ helped to: (i) translate the state programme to the organisational level; and (ii) to provide a solution to the management problem of distributive fairness (rather than, as above, efficiency).

These two cases suggest that political programmes are adopted and administered by the university one by one, i.e., the production process is modified not in a comprehensive way, but each political initiative remains isolated also at the university level. Three aspects are particularly interesting: First, the organisational autonomy is reinforced by the university's management of the inherent tensions between the political programmes, their underlying objectives, and their locally feasible implementation. Second, the university is able to provide more efficiency, effectiveness, and fairness at the same time, just not within the same programme. This increase in differentiation, that was claimed by organisational sociology (e.g., Luhmann, 1964), requires the strengthening of the organisational core. This effect can also be observed when the university launched its own 'special support programme' that aimed to (moderately) fence off the effects of political programmes and help staff to obtain funds—for example, by covering travel costs or the support of publications—which would otherwise slip between the cracks in other allocative mechanisms. This 'special support programme' is a way in which the organisation can individually manage unwanted but politically intended effects of the overall allocation regimes that interfere with the normal operations of the university. Third, the university applied these political strategies independent of each other, not because they were not able to bundle them, but because the political origin of these programmes required both public visibility (cf. Luhmann, 1981) and the decoupling of programmes. This last also enabled the university to respond more flexibly and to compensate some of the organisationally dysfunctional effects of other programmes. The university reform can be best illustrated as a *grid*, with three horizontal layers of politics, organisation, and staff, each re-organised distinctly by the vertical programmes, mostly initiated by the state (see Table 1). While a unifying strategy had been chosen at Caterpillar, the MU kept those political strategies loosely coupled, and 'invented' a variety of performance addresses: Individuals and Chairs, professional groups, and organisational substructure, such as institutes or departments, were to be held responsible for performance. This new flexibility in allocating responsibilities also enabled new management options: For example, when modules became the key addressees for performance, the university was able to (and in some cases, does) reorganise its study programmes by rearranging modules without considering disciplinary structures. The university is able to gen-

Table 1. Summarising the main features of the MU reform.

		Three Levels of Analysis		
		Political expectations	Organisational adaptation	Setting up responsible workers
How MU responds to political programmes	Performance agreements	Direct demands	Facilitator, pure administration	Department, institute, Chair
	LOM	Reward as principle	Optimise Universities performance	Chairs are held responsible
	Student funding	Substitute student fees	VSÄ as internal allocative mechanism	Professional groups around module
	Special support programmes	—	Negotiated balance and peace	Individual Researcher

erate numerous (interdisciplinary) degree courses from a small number of modules.

3.3. New Professional Roles

Quantification, i.e., the allocative procedures, together with the adaptations of the production process, demands the individual member to perform flexibly, cooperate with changing module partners, and behave entrepreneurially (also vis-à-vis the university administration, e.g., Scholz & Stein, 2010). Our interviewees, however, did not praise this new flexibility and adaptability of the academic profession, but rather emphasised that individual professors overreacted to numbers, and gamed them. For example, the interviewees were puzzled by the fact that performance indicators had such a strong steering effect given that they only covered an insignificant part of the budget. They explained this effect by the psychological disposition of professors. The examples they referred to in the course of the interview ranged from a professor 'freaking out' because he 'lost' €12, to staff systematically manipulating their performance measurements, for example by slicing one exam into three (thus, tripling their internal LOM score), or using ISBN for teaching slides and counting them as publications. These complaints suggest that professors responding disproportionally to changes in allocation are 'system optimisers' and are considered morally deficient. An 'economic' explanation would point to the fact that even the smallest differentiation in allocation may establish new social positions derived from the 'demonstrative consumption' that is made possible by these additional resources and reshape professorial behaviour (e.g., Veblen, 2009). This economic explanation can be supported by organisational studies that perceive the apparent overreaction as a reasoned response when repeated decisions of performance-based budgeting stress how critical those indicators are for their success in the organisation. By these decisions, the academic teachers are—repeatedly—differentiated as either successful/affluent members of the profession or less successful/poor ones, and they learn that even subtle differences highlight the different levels of appreciation by the organisational management, and therefore provide strong incentives for them to change their behaviour (see also Luhmann, 1981, p. 96). Thus, what the university management perceives an overreaction reflects the emergence of new organisational and political norms. Overreaction is not a personal deficiency, but rather signals the presence of tensions regarding the professional understanding of roles and the organisation. Given that only a small number of professionals overreacted, it might be concluded that the majority are well integrated into the new regime. A "new academic citizenship" emerged which strongly responds to organisational signals and the competitive structures of the regime (see also Scholz & Stein, 2010).

4. Analysis and Concluding Remarks

Although this section must start by cautioning against generalising our findings as they are based on a single case, it is also true that the layered model of Miller and O'Leary proved helpful for structuring and explaining the performance-oriented reforms at MU, and thus turned Caterpillar and the MU into comparable cases. With the university case also showing that organisational reform is not driven by organisational dynamics alone but is complemented and sometimes contradicted by political programmes, and the necessary re-socialisation of university members (see also Schiebel, 2019), the results of this study seem to have the potential to inform more comprehensive future studies in the field of higher education.

A first finding that could guide such studies confirms that politics played a critical role for the reforms. However, the MU case differs from Caterpillar as politics not only emphasised the urgency of reform but was also a critical source of change. The state introduced the idea of 'rewarding performance' in the form of performance indicators to differentiate the allocation of resources. Due to its restricted volume however, the LOM did not

fundamentally redirect the university's financial governance, rather, it operated on a symbolic level, but was nevertheless effective. This restricted change indicates the key challenge to the university management: On one hand, it takes up new responsibilities and redesigns the production process; and on the other hand, it perpetuates the university's state-dependency and its professional traditions (see Brunsson & Sahlin-Andersson, 2000). This trade-off has penetrated the very core of the reform and leaves the university with the herculean task of 'changing without change.' Politics agreed on the need to act and provided a set of programmes. However, related changes—for functionality reasons—cannot be translated directly to the university level but are partially absorbed by the organisation. The questions that follow from this observation concern the degree to which political changes reach the organisational level, if they can remain purely symbolic or to what extent do they redefine the functionality of the university?

A second finding concerns the organisational responses to resource differentiation. The performance-indicator-based strategies were—as the communication theory of quantification suggested—translated by the university into local schemes. To meet organisational needs, the MU modified performance indicators to 'correct' the effects of the LOM that were found to be counterproductive. A comparative view would enlighten us about the structural limits to such translations. As organisational sociology predicted, the diverse state programmes will lead to an internal differentiation of organisational management. The fact that the state's programmes were not bundled at the organisational level but implemented independently, can partly be explained by a political logic of visibility, but also by the fact that isolated ways to 'digest' the political strategies strengthen the organisational core and thus the MU's actorhood. The political programmes are compartmentalised by the university, not necessarily to ease or optimise the production process, but to reflect on the overall organisational purpose, management demands, and coordination problems (cf. Luhmann, 1964, p. 144). A strong indicator of this reclaimed autonomy of the administrative core is the *special programme* that should repair allocative unfairness. At the same time, such compartmentalisation provided operational flexibility and allowed the university to navigate the conflict between goals—e.g., between efficiency and fairness—and as a result, contributed to the new actorhood of the university. But this actorhood develops under conditions of state-dependency and raises the question of how far such independence is politically or organisationally feasible. This becomes tangible when contrasting the variance of political programmes and the organisational response with Caterpillar's uniform approach to simplify and modularise the production process. Some of the local solutions indicate the emergence of a strong bureaucracy. Also, the case of substituting student fees points to the strengthening of managerial functions, and it also provides the university with the possibility of flexibly combining modules and thus establishing new courses and study programmes at low cost. The production process may be simplified at the level of the individual programmes as, overall, the reform generated a confusing multitude of new addresses and competing procedures. This variety of new institutional addresses made the reform process less stable and, at the same time, allowed the university management to balance unwanted effects and reduce the biases implicit in the strategies.

Third, the organisational reform challenges traditional responsibilities and by that also the traditional roles of academic professionals. Caterpillar needed to generate a workforce that was able to read and willing to be guided by numbers; staff who should 'own' the production process. The higher education literature indicates that professions are already 'owners' of the production process and that they could (until now) resist organisational attempts to appropriate teaching and research. This assumption made the observation of the uncertainties of professorial behaviour even more interesting. The challenges to professional roles have internally been interpreted as individual and highly irritating strategies; 'system optimisers' and their 'morally deficient behaviour' indicated that new allocation strategies tend to trigger the optimisation of income rather than of output or professional quality. However, the staff's deviant behaviour showed that the new governance rules had started to have an effect, even if the university management apparently would have preferred less adaptation. The balance between steering effects and professional independence needs to be reassigned, mainly for the purpose of university management. How this could be done should be studied in greater detail.

The MU case confirms the importance of quantification for university reform, but also raises numerous new questions that require more comprehensive studies. For example, the interviewees suggested that the translation of political programmes varies by university. If the state programmes were dealt with independently, these specific local solutions would require further, comparative attention. Moreover, organisational sociology suggests that the political expectations would lead to internal differentiation and a strengthening of the central administration. Does this assumption hold for all political programmes and organisations equally or do we observe a more centralised, bundled steering of financial allocations? Furthermore, if the professional role changes, what would it mean for the profession of academics? Are professionals able to deal with the various demands that each translation process brings about or do professional roles diversify and further specialise? Also, up to what point can the professional core of the university be adapted and when does change challenge the functioning of the university?

This article placed itself between two positions characteristic of higher education studies. On one hand, a vast literature has emerged discussing the deficiencies

of universities which can be healed only by 'turning the university into an organisational actor' (e.g., Krücken & Meier, 2006), while on the other hand, the particularity of the university is highlighted (e.g., Musselin, 2007). This study emphasised the flexibility of the organisational form of the university and its ability to make trade-offs between the necessary adaptation to financial and political conditions of research and teaching, and its academic, professional core. Although this seems difficult at times, the management of and through numbers by the MU may serve as a reminder of the inevitability of the universities having to protect professionals against political, economic, or public interventions, and of the assurance that it can to a large extent be incorporated into the traditional organisational form of the university.

Acknowledgments

I gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG) (Project No. 627097) under the Open Research Area Scheme (Project Title: *QUAD—Quantification, Administrative Capacity and Democracy*). The article has greatly benefited from the discussion with workshop participants and the anonymous reviewers.

Conflict of Interests

The author declares no conflict of interests.

References

Austin, J. (1975). *How to do things with words*. Cambridge, MA: Harvard University Press.

Brunsson, N., & Sahlin-Andersson, K. (2000). Constructing organizations: The example of public sector reform. *Organization Studies*, *21*(4), 721–746

Clark, B. (1986). *The higher education system: Academic organization in cross-national perspective*. Berkeley, CA: University of California Press.

Cohen, M., March, J., & Olsen, J. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, *17*(1), 1–25.

Espeland, W., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, *113*(1), 1–40.

Espeland, W., & Stevens, M. (2008). The sociology of quantification. *European Journal of Sociology*, *49*(3), 401–436.

Heintz, B. (2012). Welterzeugung durch Zahlen: Modelle politischer Differenzierung in internationalen Statistiken, 1948–2010 [Generating the world through numbers: Models of political differentiation in international statistics, 1948–2010]. *Soziale Systeme*, *18*(1/2), 7–39.

Heintz, B. (2016). „Wir leben im Zeitalter der Vergleichung." Perspektiven einer Soziologie des Vergleichs ["We live in an era of comparison." Perspectives of a sociology of comparison]. *Zeitschrift für Soziologie*, *45*(5), 305–323.

Heintz, B. (2018). Von der Allmacht der Zahlen und der Allgegenwart des Bewertens [About the omnipotence of numbers and the omnipresence of assessments]. *Soziologische Revue*, *41*(4), 629–642.

Herz, A., Peters, L., & Truschkat, I. (2015). How to do qualitative structural analysis: The qualitative interpretation of network maps and narrative interviews. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *16*(1). http://dx.doi.org/10.17169/fqs-16.1.2092

Hillebrandt, M. (2020). Keeping one's shiny Mercedes in the garage: Why higher education quantification never really took off in Germany. *Politics and Governance*, *8*(2), 48–57.

Hood, C. (1991). A public management for all seasons? *Public Administration*, *69*(1), 3–19.

Huber, M., & Hillebrandt, M. (2019). "Pay for promise" in higher education: The influence of NPM on resource allocation in German universities. *Historical Social Research*, *44*(2), 247–269.

Krücken, G., & Meier, F. (2006). Turning the university into an organizational actor. In G. S. Drori, J. W. Meyer, & H. Hwang (Eds.), *Globalization and organization: World society and organizational change* (pp. 241–257). Oxford: Oxford University Press.

Krüger, A. K. (2020). Quantification 2.0? Bibliometric infrastructures in academic evaluation. *Politics and Governance*, *8*(2), 58–67.

Luhmann, N. (1964). *Funktionen und Folgen formaler Organisation* [Functions and consequences of formal organisations]. Berlin: Duncker & Humblot.

Luhmann, N. (1981). *Politische Theorie im Wohlfahrtsstaat* [*Political theory in the welfare state*]. Munich: Günter Olzog Verlag.

Luhmann, N. (1992). Zwei Quellen der Bürokratisierung in Hochschulen [Two sources of bureaucratization in universities]. In A. Kieserling (Ed.), *Universität als Milieu* [University as ambience] (pp. 74–79). Bielefeld: Haux.

Luhmann, N. (1997). *Die Gesellschaft der Gesellschaft* [The society of society]. Frankfurt: Suhrkamp.

Mennicken, A., & Espeland, W. (2019). What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, *45*, 223–245.

Miller, P. (2001). Governing by numbers: Why calculative practices matter. *Social Research*, *68*(2), 379–396.

Miller, P. (2007). Wie und warum das Rechnungswesen in der Soziologie in Vergessenheit geriet [How and why sociology forgot accounting]. In A. Mennicken & H. Vollmer (Eds.), *Zahlenwerk: Kalkulation, Organisation und Gesellschaft* [Creating numbers: Calculation, organization and society] (pp. 19–42). Wiesbaden: VS Verlag.

Miller, P., & O'Leary, T. (1994a). Accounting, "economic citizenship" and the spatial reordering of manufac-

ture. *Accounting, Organizations and Society*, *19*(1), 15–43.

Miller, P., & O'Leary, T. (1994b). The factory as laboratory. *Science in Context*, *7*(3), 469–496.

Musselin, C. (2007). Are universities specific organizations? In G. Krücken, A. Kosmützky, & M. Torka (Eds.), *Towards a multiversity? Universities between global trends and national traditions* (pp. 63–86). Bielefeld: Transcript.

Porter, T. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Power, M. (2007). *Organized uncertainty: Designing a world of risk management*. Oxford: Oxford University Press.

Ringel, L., Brankovic, J., & Werron, T. (2020). The organizational engine of rankings: Connecting "new" and "old" institutionalism. *Politics and Governance*, *8*(2), 36–47.

Rose, N. (1991). Governing by numbers: Figuring out democracy. *Accounting, Organizations and Society*, *16*(7), 673–692.

Schiebel, F. (2019). *Entscheidungsprobleme in Universitäten* [Decision problems at Universities] (Unpublished Masters dissertation). Bielefeld University, Bielefeld, Germany.

Schimank, U. (2005). 'New public management' and the academic profession: Reflections on the German situation. *Minerva*, *43*(4), 361–376.

Scholz, C., & Stein, V. (2010). Bilder von Universitäten—Ein transaktionsanalytisch-agenturtheoretischer Ansatz [Images of universities—A transaction-cost analytical, agency theoretical approach]. *Betriebswirtschaftliche Forschung und Praxis*, *62*(2), 129–149.

Stichweh, R. (1994). Die Form der Universität [The form of the university]. In R. Stichweh (Ed.), *Wissenschaft, Universität, Professionen* [Science, university, professions] (pp. 246–277). Frankfurt: Suhrkamp.

Veblen, T. (2009). *The theory of leisure class*. Oxford: Oxford University Press.

Weber, M. (1978). *Economy and society: An outline of interpretive sociology*. Berkeley, CA: University of California Press.

Wittrock, B. (1993). The modern university: The three transformations. In S. Rothblatt & B. Wittrock (Eds.), *The European and American university since 1800: Historical and sociological essays* (pp. 303–362). Cambridge: Cambridge University Press.

About the Author

Michael Huber is Professor of Sociology at the Department of Sociology at Bielefeld University. His research is focused on organisational sociology, higher education studies, and regulatory policymaking, more specifically, risk-based regulation.

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2020, Volume 8, Issue 2, Pages 36–47
DOI: 10.17645/pag.v8i2.2576

Article

The Organizational Engine of Rankings: Connecting "New" and "Old'' Institutionalism

Leopold Ringel *, Jelena Brankovic and Tobias Werron

Faculty of Sociology, Bielefeld University, 33615 Bielefeld, Germany; E-Mails: leopold.ringel@uni-bielefeld.de (L.R.), jelena.brankovic@uni-bielefeld.de (J.B.), tobias.werron@uni-bielefeld.de (T.W.)

* Corresponding author

Submitted: 22 October 2019 | Accepted: 19 December 2019 | Published: 9 April 2020

Abstract
When explaining the ubiquity of rankings, researchers tend to emphasize macro or contextual phenomena, such as the power of or the trust in numbers, neoliberal forces, or a general spirit of competition. Meanwhile, the properties of rankers are rarely, if at all, taken into account. In contrast to the received wisdom, we argue that the institutionalization of rankings in different fields is also contingent upon another, often-neglected factor: Over time, rankers have become increasingly more organized. To investigate the role of ranking organizations, we look into the distinct properties of present-day rankings and highlight three dimensions along which rankings have evolved over the course of the twentieth century, namely, publication frequency, handling complex tasks, and audience engagement. On this basis, we argue that these dimensions have to a large extent been affected by formal organization and we show how ranking organizations have over time developed capacities to: (a) publish rankings on a continual basis; (b) handle the often complex production process by means of division of labor; and (c) generate considerable degrees of attention by addressing large and diverse audiences. On a more general note, we argue that accounting for the role of organization in the instutionalization of rankings requires a combination of insights from both "old" and "new" strands of thinking in institutional theory.

Keywords
audiences; institutions; organizations; quantification; rankings

Issue
This article is part of the issue "Quantifying Higher Education: Governing Universities and Academics by Numbers" edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

1. Introduction

Higher education, sports, business, human development, and the arts are only some among the growing number of areas which are nowadays routinely subjected to novel forms of quantified comparisons and monitoring. Among these, rankings stand out as particularly pervasive and impactful. Be it the Human Development Index, which the United Nations Development Programme uses to measure the development of nations, TripAdvisor's Popularity Index listing the best restaurants and hotels in a city, or the Times Higher Education's (THE) World University Rankings comparing universities on a worldwide scale—rankings, it seems, play an increasingly important role in the contemporary world.

The extant literature, however, rarely addresses the problem of *why* rankings have become so pervasive. Instead, the effectiveness of rankings is often taken for granted, while little is added to the explanation of the social processes undergirding their institutionalization. Rather, the literature tends to refer to broader trends in society which rankings are part of, such as the onset of digitalization (Mau, 2019), the aura of rationality surrounding numbers (Espeland & Stevens, 2008), the institutionalized trust in numbers (Porter, 1996), the growing importance of performance measurement in governance

(Mehrpouya & Samiolo, 2016), neoliberalism (Münch, 2014), or even a general spirit of competition in mankind (de Rijcke, Wallenburg, Wouters, & Bal, 2016).

Following our earlier work, we see the institutionalization of rankings as intimately connected to discursive processes unfolding historically in different fields (Brankovic, Ringel, & Werron, 2018; Ringel & Werron, in press; Werron & Ringel, 2017). In plain terms, as rankings become increasingly more present in public discourse, be it as objects of praise, criticism or simply "neutral" reporting, they are more likely to become institutionalized. An integral part of this process is that rankings keep the ranked "on their toes": they create pressure for those who are subject to them to strive to not lose their position, maintain it, or improve it (Espeland & Sauder, 2016; Esposito & Stark, 2019). However, we know only little about if and how rankers themselves add to this, given that they are routinely overlooked in the extant literature (Rindova, Martins, Srinivas, & Chandler, 2018). So far, we have learned only little about the properties, practices, and strategies devised by producers of rankings and how they intersect with or add to the discursive institutionalization.

What is then to be said about those who produce rankings? While for a long time the process has predominantly been undertaken by individuals who acted in their personal capacity as, for example, scientists or critics, and who typically had only marginal organizational support, contemporary rankings are usually produced by organizations of various kinds, such as for-profit businesses, newspapers, international governmental and non-governmental organizations, or universities. A closer look at some of the well-known inventories of rankings and similar devices (e.g., Bandura, 2011; Kelley & Simmons, 2014) reveals that the most "successful"—that is, regularly published, well-known, and impactful—rankings today are the product of organizational efforts and were first published usually between the 1980s and early-to-mid 2000s.

To address this, in this article we examine the (largely neglected) role of organizations in the rise and discursive institutionalization of rankings. Broadly speaking, when it comes to the pervasiveness of organizations in modern society, two explanations have dominated organizational scholarship throughout the twentieth century (Selznick, 1996; Stinchcombe, 1997). One body of literature in organizational institutionalism, that is, the so-called "old" one, stresses the capabilities of formal organizations to complete complex and resource-intensive tasks as opposed to other forms of social organization. In contrast, the new institutionalist perspective, which emerged partly in response to the old institutionalist tradition, explains the explosion in numbers of organizations populating the globe by emphasizing that formal organization itself has become a cultural institution granting legitimacy to those embracing it (Meyer & Bromley, 2013; Meyer & Rowan, 1977). In simplified terms, formal organizing thus abounds either because it has proven effective or because it is legitimate. Although, by and large, the legitimacy argument is today considered a superior approach to this old question, we argue that revisiting the tenets of old institutionalism holds promise for exposing some overlooked elements of the rankings phenomenon.

With these two potential trajectories of explanation in mind, in this article we review empirical studies focusing on the producers of rankings in a diverse set of fields, particularly, in the arts, science and higher education, education, tourism, and business. Specifically: in the first step we (a) scan the literature on rankings to identify potential relations between the organizational expansion and the rise of rankings; and, in the second step, we (b) look for evidence of the effectiveness argument, rather than that of cultural legitimacy, in the role organizations play in the institutionalization of rankings. By unveiling the properties, practices, and strategies devised by key actors in discursive arenas, the producers of rankings, we wish to open a new avenue in the broader research on their global institutionalization (Brankovic et al., 2018; Ringel & Werron, in press).

2. The Rise of Organizations: New Institutionalism and the "Old"

Since the nineteenth century and particularly since the end of World War II, we are witnessing a rampant spread of organizations across sectors of society (Drori, Meyer, & Hwang, 2006; Meyer & Bromley, 2013). The new institutional theory, in the tradition of Meyer, Rowan, Powell, DiMaggio, and others (Meyer & Rowan, 1977; Powell & DiMaggio, 1991), argues that formal organization as such has become a cultural institution embodying and symbolizing the promise of our highest ideals of rationality. The often described "rationalization" of society over the past decades, that is, the spread of universalized and abstract cultural beliefs and templates about equality, education, human rights, gender and so forth, thus "creates a framework that encourages organizing in a wide range of societies and domains" (Meyer & Bromley, 2013, p. 367) beyond actual functional necessities. These beliefs and templates are to a large degree theorized and legitimized by science—a process that new institutional theory refers to as "scientization" (Drori et al., 2006). Modern world society is therefore made up of a dense latticework of scientized norms, values, standards, beliefs, and ideas, all of which are connected to specific visions of organizational actorhood. Accordingly, studies in this tradition have shown that organizations are often more busy incorporating such structures so as to adhere to so-called "rationalized myths" in order to be granted legitimacy by their environment.

How do scientized templates about formal organization and organizational governance find their way to the public and, by extension, trigger the creation of, or change in existing organizations? How are they, as institutional theory would have it, diffused? The literature

suggests two sets of answers. One builds upon the work of DiMaggio and Powell (1983) which conceives of society as comprising a multitude of organizational fields and outlines three isomorphic processes of diffusing organizational templates within fields: coercive (organizational structures are diffused via laws or otherwise formally binding rules based upon sanctions); normative (organizational structures are diffused based upon the belief that they are "right"); and mimetic (organizational structures are diffused because they are taken for granted). The other draws on Strang and Meyer (1993) who emphasize that cultural templates often do not travel directly through communicative networks, but are diffused indirectly, for instance, through observation. They argue that cultural templates are more likely to be spread when they are theorized: "By theorization we mean the self-conscious development and specification of abstract categories and the formulation of patterned relationships such as chains of cause and effect" (Strang & Meyer, 1993, p. 492). Building upon these insights, Werron (2014) suggests defining observers who create such theorizations as "universalized third parties." Such third parties, of which the producers of rankings are an example, specialize in publicly observing actors and addressing audiences, thereby indirectly creating pressure among the said actors to follow the dominant models.

Capabilities of formal organizations, therefore, do not play a major role in the institutionalization of formal organization. According to the new institutionalist line of argument and evidence, formal organizations, put simply, do not necessarily exist in such great numbers because they are (more) effective (than other forms of social organization). If and under which circumstances organizations are or need to be effective in carrying out certain tasks, is then a completely different and, from the perspective of this tradition, secondary question, or at least one that it empirically rarely explores.

In the large body of literature which attributes the rise of formal organization to its superior capabilities to achieve certain tasks, for instance when it comes to administering a region or a country, producing goods, and taking care of the sick, three classics are worth revisiting here. Weber (1915/1947) famously defined organizations as beacons of western rationality due to their unique properties: They produce written rules of conduct, outline specific spheres of competence, divide labor, have clear-cut authority structures, employ trained personnel to fill positions, have some degree of discretion in resource allocation, and have members who typically do not own the means of production but instead receive wages. Barnard (1938) argued that creating formal organizations becomes a necessity when individuals face too many goals and tasks which are too complex for them to accomplish on their own. He also stressed the importance of informality as a complementary mode of action and mechanism to increase support among employees. Coase (1937) was also interested in question as to why individuals feel the necessity to create formal organizations. He highlighted the transaction costs one has to carry when working towards achieving a goal and emphasized that there are essentially two options: Services can either be bought on a market or made. In complex situations, for example, when frequent discussions, deliberations, and negotiations are necessary, or, when products have to be constantly refined and repaired, seeking vertical integration by means of creating a formal organization is more rational than buying these services on a market.

In the following decades, a vast body of scholarship critically engaged with these contributions. Nevertheless, they in principle maintained that formal organizations were a modern institution that, for all their flaws, arguably still did some things more effectively than any other form of social organizing and is for this reason an all-pervasive feature of modern society. In public administration studies, for instance, Simon (1945) notably argued that due to anthropological cognitive constraints, individuals exceed their analytical capabilities quickly and thus need to be able to cooperate with others in order to compensate for their imperfections. As a result, organizations are characterized by "bounded rationality": They do act rationally only to a certain degree—and more so than any other type of collective actor (Simon, 1957). This, Simon explains, is because organizations allow individuals to specialize (for example, by observing the environment only in legalistic terms as a company lawyer), which allows them to put their limited cognitive abilities to better use than if they had to consider a multiplicity of aspects (we could, for example, imagine a lawyer who also had to be a PR manager and an expert in the production of goods).

The first and second generation of American organizational sociologists, among which are Robert K. Merton, Philipp Selznick, and Alvin Gouldner, were particularly interested in revisiting Weber's claim of the superiority of organizations by studying intra-organizational processes. They unveiled a variety of unintended and sometimes dysfunctional consequences. Nonetheless, as Selznick (1996) and Stinchcombe (1997) argued, they as well acknowledge that organizations are, to some degree, more efficient and effective at accomplishing the tasks at hand. They are able to do so by devising formal rules, which can broadly be defined as "abstractions that govern" (Stinchcombe, 2001, p. 43). Furthermore, Stinchcombe makes the case that institutions are not self-supporting, emergent structures with a life of their own but in need of support by actors on the ground. Institutions, in short, can only survive if organizations deliver services promised by the institution and if the quality of these services is considered acceptable.

Borrowing from the title of Stinchcombe's book (2001), we might summarize the thrust of this body of literature as follows: There are indeed times "when formality works." There is thus a rich variety of contributions to organizational research arguing for conceptualizing formal organizations as a modern phenomenon

that has become institutionalized because it does—at least some—things more effectively than any other form of social organization. Formal organizations divide labor, standardize processes, and employ trained specialists (Weber); they coordinate human behavior while retaining autonomy in setting and changing goals (Barnard); they deliver service which could not be bought as easily on a market (Coase); they compensate the limited cognitive capacities of individuals (Simon); and they possess the ability to employ distinct mechanisms of governance (Stinchcombe).

Summing up, according to new institutionalism, we would expect the institutionalization of rankings to lean heavily on the *cultural institutionalization* of formal organization. The "old" institutionalism, on the other hand, would see the institutionalization of rankings as primarily contingent upon the *capabilities* of formal organizations, which are in their effect superior to other forms of social organizing. In the following, rather than seeing the two brands as mutually exclusive, we use their insights as heuristics to analyze the role organizations play in the institutionalization of rankings. In the process, we wish to show that institutionalism could in fact be practiced as a way of theorizing that incorporates insights from both the "old" and the "new" tradition.

3. The Expansion of Rankings along Three Dimensions

To examine the role organizations play in the institutionalization of rankings, we reviewed the empirical literature on rankings with a specific focus on studies examining their producers. Against this backdrop, we then selectively searched for first-hand information on much discussed and long-lasting rankings in different fields, using publicly available data (published reports, websites, and secondary sources). We systematically compared the cases and organized our insights into several themes. Discussing these insights, this section specifies a variety of ways in which organizations matter in making rankings a regular feature of many fields in modern society. The first dimension refers to the fact that rankings usually address and simultaneously *maintain a problem*, thus inducing their regular and continuous publication as opposed to some of the earlier experiments with rankings which tended to be one-off efforts. The second dimension refers to matters of *handling complex tasks*, that is, an evolution from the production of relatively simple lists, simple indicators, and ranking a relatively small number of entities, to more complex rankings, based on composite indicators and with a larger span. The third dimension, *audience engagement*, refers to a transition from relatively narrow expert audiences to larger, more diverse, non-expert or lay ones. We take each of the three dimensions in turn and selectively draw from examples of rankings that were produced before the ranking-frenzy of the last thirty years, contrasting them with more recent rankings in which organizations play more significant roles.

3.1. Addressing, Producing, and Maintaining the Problem

Rankings neither appear nor exist in a social vacuum. They frequently address societal problems that are in many cases perceived as major or even global challenges, such as education, corruption, or climate change. Even national rankings usually refer to issues that are of universal nature and typically do not touch upon matters that would only make sense in a local context. On the contrary, they normally draw from (global) templates offered by "universalized third parties" (Werron, 2014)—particularly international (governmental or non-governmental) organizations and universities. We might therefore suspect that, following the widespread theorization of global challenges in the past decades, rankers should find it relatively easy to draw from existing templates in order to define a problem to which their product should be an answer or a road to one.

However, rankings do not only address existing issues, but at times they also actively work towards creating the very problem to which, they claim, their product is the remedy. In our earlier work on global university rankings, for example, we demonstrated that, although such rankings nominally intend to measure something that is supposedly "out there," in doing so they discursively construct a distinct notion of scientific "excellence" (Brankovic et al., 2018). In many ways, rankers such as the Shanghai Ranking or the THE World University Rankings do not just map the existing global field of universities, but instead they importantly contribute to its emergence as a shared social space.

Whether rankings address existing challenges or actively create new ones, in order for them to survive as rankers, they have to *maintain* the "problem." This, we argue, is contingent upon their ability to keep the operation going. A crucial prerequisite of doing so is, logically, to secure funding. However, as rankings are produced by a host of different types of organizations, there is no one clear path for all of them to follow. For-profit organizations, such as the U.S. News or Mercer make rankings part of their business strategy (for higher education check, for example, Stack, 2016). Many rankings, on the other hand, are produced by public and non-profit organizations such the OECD, Transparency International, or the World Bank, and depend upon an often complex and dense network of funding bodies, ranging from government agencies, prominent international organizations, private foundations and philanthropies, to corporations.

Yet, however ripe the cultural and symbolic infrastructure for the birth of ever new rankings, for rankings to become effective they also have to be published repeatedly and sometimes even regularly (Brankovic et al., 2018; Werron & Ringel, 2017). Surprisingly, research often does not take note of the fact that present-day rankings are typically produced weekly, monthly, annually, or biannually. To pinpoint the struggles such continual publication creates, it is illuminating to contrast modern-

day rankings with their predecessors (Brankovic, Ringel, & Werron, 2019). In fact, many of the early rankings and similar forms of quantitative evaluation (that is, quantified comparisons sharing some, though not all properties of modern rankings), published in the pre-1980s area, were one-off experiments by individual experts in their respective fields who had no or, at best, minor organizational support.

Take the art field for instance: Even though efforts to establish numerical forms of evaluation in the arts can be traced back to the eighteenth century, they only have become successful since the 1970s (when the periodic art ranking, the German *Kunstkompass*, was first published) and particularly in the 2000s (Buckermann, 2019). Spoerhase's (2018) study on art rankings in the eighteenth and early nineteenth centuries illustrates this point. He traces quantified comparisons of (mostly) deceased artists produced by other artists and art critics, starting with the publication of Roger de Piles' *Balance des Peintres* in 1708 and ending with Jean-Francois Sobry's *Balance de Peintres, Rectifée* in 1810. In a similar exercise about a century later, surrealist rankings published in the journal *Littérature*, edited by surrealist artists André Breton, Lois Aragon, and Philippe Soupault, were also created as singular experiments (Schmidt-Burkhardt, 2005).

Science and higher education provide another instructive example. A review of higher education rankings between 1900 and 1980 reveals that almost all of them had only been published once (Webster, 1986). Among the first to create such league tables was the psychologist James Cattell, professor at Columbia University and editor of the journal *Science* (Hammarfelt, de Rijcke, & Wouters, 2017; Ringel & Werron, in press), who produced a variety of one-off rankings and whose method was copied by many of his successors. Even in cases when a ranking exercise was repeated, such as those by Raymond Hughes, published in 1925 and 1934, and the so-called *Cartter Report*, published in 1966 and 1970 (Webster, 1986), the periods between the publication of the first and the second report—nine years and four years, respectively—were relatively long and certainly cannot be considered regular.

What are the reasons for the absence of repetition in the production of these rankings, both in the arts and in the scientific field? A tentative answer could be their purpose. Spoerhase (2018) suggests that the rankings and similar devices he studied did not seem to aim at measuring current performances, to be re-assessed at a later stage, but rather resembled what we today sometimes refer to as "best of all times" lists. After publishing such lists, individual rankers lacked interest in furthering their assessments. In the scientific field, the previously mentioned Cattell produced rankings as part of his studies on the origins of "eminence," as a relatively stable category and he did not seem to show much interest in capturing change as such. Organizations, on the other hand, have means to "force" their members to have a long-term interest in rankings: Within a "zone of indifference" (Barnard, 1938), members of an organization can be expected to do as they are told irrespective of their emotions, preferences, or (changing) interests.

A second reason for the one-off nature of many of these early rankings in different fields seems to be that repeating them would have been not only costly, but also time-consuming: Producing anything on a regular basis instead of once obviously has clear implications for the workload to carry and the money to spend—especially in the case of individuals for whom producing rankings is not their sole or even their primary responsibility. Hammarfelt et al. (2017), for example, detail the amount of time and energy Cattell had to dedicate to produce one ranking. He first had to scan encyclopedias, make a list of "eminent" scientists in different fields, prepare a questionnaire, select those who should receive it, send out letters to them with specific instructions, wait for the letter to return, weigh scores, calculate overall results, and, finally, publish his findings on the "scientific strength" of universities in the journal *Science*, which he only could because he himself was the editor. The authors suggest that Cattell did not repeat this exercise due to a "lack of data." They further speculate that as the biographical directory *American Men of Science*, which was the foundation of his sample, grew in size, Cattell on his own was not able to maintain the "calculative power" necessary for compiling and reproducing the ranking. They conclude that "(i)t may have been too much for Cattell" (Hammarfelt et al., 2017, p. 405).

All of this stands in stark contrast with how today's rankers are able to handle their workload. Besides their superior ability to retain financial resources mentioned earlier, they often have trained personnel specialized in and paid for the production of rankings. In their study of rankings in the IT sector, for instance, Pollock and D'Adderio (2012) showcase how contemporary ranking organizations such as Gartner have an "army" of employees who are responsible for the creation and maintenance of so-called "magic quadrants." This indicates that in the case of many contemporary rankings, the employment of those responsible for the production process is contingent upon the ranking exercise to be continued. In turn, if individuals decide to leave the organization, they can be replaced by others who continue their work. Take the Corruption Perception Index: Even after its well-known founder, Johann Graf Lambsdorff, retired, and wanted to retire the Index (Global Integrity, 2009), Transparency International continued the publication.

Not only employees, but also the ranking organizations themselves may have incentives to commit to continuing the production. This is especially true when a ranking becomes so successful that the survival of the organization depends on it. When it produced its first college rankings in the early 1980s, U.S. News, a media company, was in financial trouble. The publication of its first successful—and still existing and widely read—ranking practically saved the business. In the coming

years, U.S. News became increasingly more invested in the production of rankings (Espeland & Sauder, 2016). In 1990, less than a decade after its first publication, U.S. News expanded its ranking to include all four-year colleges in the U.S., alongside its traditional top 25 lists (Jin & Whalley, 2007). Pigeon Paradise, an auction house producing league tables of pigeon races, is another example of an organization rapidly expanding through rankings. They started out in the 2000s as a club ranking pigeon races in Central Europe. This soon spurred interest in Asia, which attracted considerable financial resources as buyers sought the most successful pigeons. The growth which ensued led to the Pigeon Paradise proclaiming itself the "engine of the pigeon market" (Bahrami & Meyer, 2019). In these cases, two complementary characteristics of organizations seem to reinforce each other: While the *effectiveness* of organizations contributes to the regular publication of rankings, once established, the rankings use the "self-preserving tendencies" of organizations for their long-term institutionalization.

3.2. Handling Complex Tasks

Over time, rankings in different fields have arguably become more elaborate and more complex, which can add to their perceived significance and legitimacy. Ranking organizations play a crucial role in this development as they can over time become more efficient in managing the collection, quantification, evaluation, and aggregation of ever more information about an increasing number of cases. To give some examples: The Programme for International Student Assessment started out as a ranking of 43 OECD and non-OECD countries and in its last edition in 2015 included 72 countries; the Corruption Perception Index by Transparency International went from 41 countries in 1995 to 180 countries and territories in 2018; and the website ArtFacts, founded in 2001, constantly increases its sample, currently holding at a staggering number of 601,331 artists. However, organizations do not only increase the sample of individual rankings, they also often multiply the population. The U.S. News, on the other hand, did not merely expand its original ranking by including more universities, but also began to "export" rankings into other domains, such as hospitals, law firms, vacation destinations, and even cars.

In more abstract terms, contemporary rankers often act as what Latour (1987) refers to as "centres of calculation," that is, they mobilize, translate, standardize, quantify, and evaluate information by developing large-scale networks. Due to their calculative power, formal organizations are able to include a large number of ranked entities. Such requirements greatly surpass the capacities of single individuals, or even small teams. To produce a ranking of colleges in the U.S., Kunkel (1915), for example, had to rely on his friends:

> The data upon which this paper is based were secured from three different volumes of "Who's Who," the only ones at the time available, in order that I might profit by the very kind assistance of my friends, Mr. and Mrs. Marion H. Hedges, without which I fear I would not have carried out the investigation. (p. 317)

Such solutions are today, arguably, less common. Furthermore, producing complex rankings which include a large number of entities usually requires a high degree of expertise and specialization. The personnel of ranking organizations is typically highly educated, in possession of specialized knowledge, and sometimes receives additional extensive training—thus making up a large labor force that serves as an important infrastructure for the emergence of ever more rankings. Organizations also rely on their ability to mobilize the input and cooperation of experts and academics. The Freedom House's Freedom in the World Index, for example, assembles "a team of in-house and external analysts and expert advisers from the academic, think tank, and human rights communities"; "[t]he 2020 edition involved more than 125 analysts, and 40 advisers" (Freedom House, 2020). The main task of the analysts is to evaluate their respective countries of expertise by using "a broad range of sources, including news articles, academic analyses, reports from nongovernmental organizations, individual professional contacts, and on-the-ground research" (Freedom House, 2020). The analysts then produce quantified scores, which "are discussed and defended at a series of review meetings, organized by region and attended by Freedom House staff and a panel of expert advisers" (Freedom House, 2020). Another example is the World Bank's Ease of Doing Business Index, which relies on both in-house and external expertise and in doing so consults thousands of experts who fill out the questionnaire, interacts with these individuals, and then even does spot-checks in a variety of countries (World Bank Group, 2018). Arguably, only an organization that is able to manufacture high levels of legitimacy, controls vast resources, and has extensive staffing and global networks of expertise at its disposal could accomplish a task as complex and resource-intensive.

While centers of calculation still steer the processes and, to some degree, are able change the outcomes (see Pollock & D'Adderio, 2012), other organizations create and maintain what Kornberger, Pflueger, and Mouritsen (2017) call "evaluative infrastructures"—a mix of technologies, assemblages, institutional arrangements, cultural rules, norms, habits, and conventions, allowing for the collection, creation, and procession of large amounts of data. Typical examples are platform organizations which depend on users' sharing their experience. In such cases, the organizations provide certain rules and framings (such as algorithms), but the resulting evaluations, ratings, and rankings emerge in a more or less organic fashion. TripAdvisor, for instance, has an algorithm that produces its popularity index which ranks hotels and restaurants in different areas, relying on user review (Jeacle & Carter, 2011; Orlikowski & Scott, 2013). Even

though the production of the Popularity Index proceeds almost automatically as soon as the algorithm is in place, TripAdvisor still has to provide a great deal of maintenance work. Moreover, operating at such a large scale in some sectors, such as tourism, involves setting up contracts with a variety of auxiliary organizations, engaging in PR activities, providing help desks and counseling in cases of problems, and dealing with a variety of legal questions and even lawsuits.

The increasingly complex evaluations and calculations, and not least their expansion, require a certain degree of standardization, but also distance from the object of ranking. Just like any other, ranking organizations strive to improve their management, as well as to develop strategies to socialize their members, allowing them to coordinate actions of often large, diverse, and spatially dispersed groups of people (Ashforth, Sluss, & Harrison, 2007; Maanen & Schein, 1979). Orlikowski and Scott (2013) detail the formally standardized learning process of newly hired staff at the Automobile Association which publishes a yearly accommodation guide in Great Britain. The authors describe a well-honed, elaborate, and standardized process of peer-learning at the end of which newcomers become experienced—and trusted—hotel evaluators. Similarly, in their study on the Access to Medicine Index, Mehrpouya and Samiolo show how the Access to Medicine Foundation constantly reminds its analysts to suppress their subjective impressions and feelings and analyze data as a "robot" (Mehrpouya & Samiolo, 2016, p. 21). With such practices in place, ranking organizations are not only able to standardize the behavior of their employees, but also to deal with its fluctuation. Their ability to handle ever more complex tasks, then, can be seen as both a prerequisite and an effect of their involvement in the continual production of rankings.

3.3. Engaging Large and Diverse Audience

The aforementioned examples from the arts and sciences indicate that the early producers of rankings were typically experts in the field, addressing a small circle of people and expert audiences, usually their professional peers (other artists and scientists, respectively). The compilers of art (proto)rankings, for example, even deemphasized the validity of their findings (Spoerhase, 2018), signaling that they respected institutionalized structures of connoisseurship at the expense of promoting their easy-to-understand ranking tables to mass audiences. Similarly, in the field of science, "early attempts had in common that they originated from the sciences themselves and, although they claimed to be of relevance for students, their audience mainly consisted of fellow scholars" (Hammarfelt et al., 2017, p. 406).

Since the 1980s and 1990s, more and more rankings are explicitly produced for non-expert and in other ways diverse audiences. In many cases, the very meaning of the ranking is based on the idea of transforming expert judgments into information for broader publics. However, while many accounts emphasize the numerical authority of rankings, empirical studies also highlight their aesthetic appeal. In their study on Gartner—an IT industry ranking organization—Pollock and D'Adderio (2012) describe how the creators of Magic Quadrants always keep in mind that a ranking has to deliver a "beautiful image"—a distribution that makes sense to clients, thus neither including too many nor too few cases, while offering a meaningful spread. In a similar vein, Mehrpouya and Samiolo (2016) show how ranking organizations strive to create what their informants call a "good distribution." In Latoursian terms, ranking organizations invest a great deal of time and resources to craft rankings as powerful and appealing "inscriptions" spanning and travelling a multiplicity of contexts (Latour, 1987).

In contrast to individual rankers, who seem to focus more on interacting with other experts, ranking organizations often strive to maintain interest in their product across several audiences, some of which are more and some less competent when it comes to understanding the complex and multifaceted calculations undergirding rankings. To address various audiences, ranking organizations either develop hybrid practices, that is, they engage in activities spanning multiple fields (Furnari, 2014), or divide labor processes, i.e., they set up different departments focusing on technical tasks *and* institutional environments amounting to what is often referred to as decoupling (Meyer & Rowan, 1977). Crucially, all of the practices depend on the specific capacities of formal organizations, particular division of labor in professional expertise, ranging from technical abilities (e.g., statistics, accounting) and management to communication skills (e.g., scientific, marketing, public relations).

Rankers' ability to span multiple audiences also allows them to have both the personnel specialized in adhering to a noble cause (for instance, protecting the environment, or improving healthcare) and engaging in commercial activities to generate revenue. Their complex structure enables organizations to even use their products to promote certain causes via rankings and then provide remedial services. Kornberger and Carter (2010) detail how the Anholt City Brand Index ranks global cities to promote the idea of a "city brand" and then establishes itself as a consultancy to help cities improve their brand. The aforementioned Pigeon Paradise simultaneously addresses expert audiences (so-called "pigeon fanciers") and uses its rankings to increase the valor of pigeons, which it then auctions off to the highest bidder (wealthy investors in Asia). These examples indicate that owing to their diverse and qualified workforce as well as their ability to coordinate the action of multiple departments via managerial oversight, organizations are well equipped to expand into a multiplicity of fields. With modern society comprising exponentially more fields over time (Christin, 2018; Fligstein & McAdam, 2012; Krause, 2018), such capabilities are in great demand when it comes to the handling of devices such as rankings.

Another important factor for the engagement of large(r) and (more) diverse audiences is that rankers often end up attracting attention by creating "news." In contrast, early rankings were rarely set up with this goal in mind. It is therefore little surprising that rankings which actually gained traction and were published regularly before the 1980s, such as the *Kunstkompass*, an artist ranking by German journalist Willi Bongard, actually could rely on organizations to deliver certain tasks such as marketing. In the case of the *Kunstkompass*, the journal in which it was published took care of printing, distributing, and promoting the ranking over the years, allowing Bongard to use his time to engage in the public discourse surrounding his ranking and to address criticism (Wilbers, 2019). Contemporary rankings, however, are often geared towards maximizing the attention of a large, lay and in many cases also global audience (Brankovic et al., 2018; see also Kornberger & Carter, 2010). Not only do they take great care of how they visualize their products, but they usually also employ communication or PR experts responsible for monitoring, evaluating and engaging with stakeholders. Their tasks include, among others, disseminating reports, communicating their product in a language which is more accessible to a broader audience, organizing events, and devising elaborate social media strategies.

The production of rankings that reach beyond narrow circles of experts, especially those addressing global and diverse audiences, increases the likelihood of criticism. Almost by definition, regularly published rankings are in a continuous battle for legitimacy and are often fiercely debated. The Corruption Perception Index has been accused of furthering U.S.-American interests (Gutterman, 2014), while university rankings regularly face pushback (Dörre, Lessenich, & Singe, 2013; Espeland & Sauder, 2007). Criticism and efforts to avoid it can sometimes affect the way a ranking is made and push the ranker to be "deliberately less bold" and even—as it was the case with the aforementioned Hughes' and Cartters' reports—to completely refrain from presenting the findings in a rank order (Webster, 1992, p. 252) as a way to "de-emphasize the pecking-order relationships" (Roose & Andersen, 1970, p. 2).

Some of the most prominent rankers today certainly do not shy away from controversy. To address critics and sceptics, they often devise elaborate strategies. The earlier mentioned social media activity is one such example, while organizing events is another. THE, for instance, organizes summits and launches, which they use to also get in touch with critics and/or experts, "refine" their methodology, and see how they can "do it better" next time (Lim, 2018). Organizations are in general able to counter criticism quite effectively as they have the means to orchestrate such collective responses by providing its members who appear in public with "fronts, appearances, manner, routines" (Manning, 2008, p. 680), thus allowing them to promote a favorable self-presentation (Ringel, 2018), even in spite of backlash.

Another strategy is to involve those who could potentially criticize and even de-legitimize the ranking in its production process. A classification of journals by the Danish Agency for Science, Technology and Innovation, for instance, was created by assigning the task of categorizing journals to scholars from the respective fields (Jensen, 2011). The agency thus acted as a center of calculation by branching out its calculative processes, while at the same time creating barriers against criticism: The evaluation criteria of the journals, if challenged, would be the responsibility of "the scientific community"—not of the Agency. Arguably, the ranker, which in this case was a ministry, granted legitimacy and authority to the process—something which would be far more difficult for an individual to achieve.

4. Concluding Discussion and Outlook

The expansion and effectiveness of rankings is certainly contingent upon their acceptance as a modern rationalized institution. Research in the framework of new institutionalism has time and again demonstrated that collective actors who are perceived as legitimate have easier access to resources than those who cannot or do not wish to incorporate institutionalized structures. Even more fundamentally perhaps, we have to recognize that "organizations can best communicate with other organizations," that is, organizations consider other organizations "the only adequate points of contact" (Kühl, 2015, p. 263). Hence, it is easier for ranking organizations to acquire funding and establish cooperation compared to other types of actors, such as individuals, simply because they are organizations. As a result, we might attribute the dominance of organizations among modern rankers today, to some degree at least, to the symbolic quality of formal organizations as such.

In this article we have argued that the new institutionalist thesis on the global-cultural institutionalization of "organization" could benefit from the (often forgotten) insights offered by the "old" institutionalism. We, however, do not wish to undermine cultural rationalization as an important driving force behind global organizational expansion; organizational expansion as such is, after all, not what we have tried to account for here. Rather, we call attention to the aspects of the expansion or "success" of contemporary organizational *rankers* which, we argue, cannot be accounted for exclusively though the new institutionalist lens. We neither wish to claim that rankings become institutionalized *because* they are produced by organizations; rather, we argue that organizational capabilities are an important element in the larger mechanism of this complex process. Finally, we do not claim that individuals cannot produce and reproduce popular rankings, but those are, to our knowledge, exceptions, rather than a rule.

To elaborate this point further, and drawing on the empirical evidence mentioned thus far, we suggest that organizations enable the ongoing production and pro-

motion of rankings in at least three ways. First, plainly put, compared to individuals, organizations are able to do more things at once. Transparency International, for instance, has a secretariat that is, among other things, involved in the production of the Corruption Perception Index; it also has national chapters all over the world engaging in different activities, such as advocacy, networking, engaging with the media, and lobbying. Extreme cases are large and highly influential international organizations often employing hundreds of people, with the World Bank, the OECD, and the United Nations being well-known examples. For such organizations, producing a ranking does not necessarily mean setting other tasks aside. Take the World Bank: While producing and promoting Indexes, such as Doing Business Index and Human Capital Index, it provides loans to countries, funds all kinds of projects, conducts research, and engages in advocacy.

Second, organizations can do more things at once—for a longer time. They often outlive individuals—even their founders—which makes them more likely to sustain a long-term interest in the production of rankings. Individuals, on the other hand, not only eventually pass away, but they also may lose interest in the production of rankings as they become dedicated to other endeavors. The repeatedly mentioned James Cattell is a prime example: Albeit showing great interest in mapping scientific "eminence," he was also involved in the promotion of university reform and anti-war proclamations during World War I, which eventually consumed much of his time and even put him in the position of losing his tenure at Columbia University. Put differently, there are limits to what a single person, or even a group for that matter, can do in their lifetime—a restriction that does not in principle apply to formal organizations.

Third, organizations can do more things at once, for a longer time—and can accumulate more resources, often from a greater variety of sources. Producing and maintaining a ranking (or several different rankings) is in certain cases extremely resource-intensive. Individual producers of rankings were therefore either forced to invest their own resources or wait for opportunities to arise. Cattell, for instance, had the advantage of editing a leading journal in which he could publish more or less at his own will. Kunkel, as we have seen, relied on friends. Individuals, as opposed to organizations, are also less likely to secure the funding necessary to acquire expensive instruments and technologies needed to ensure the calculative power for the production of the rankings they envision.

In conclusion, while in this article we acknowledge that the diffusion of rankings in contemporary society is largely a matter of discursive institutionalization, we wish to draw attention to the properties, practices, and strategies devised by the actors responsible for their ongoing production, which have largely remained ignored by rankings research to date. We therefore argue that formal organization is, in a way, a vital cog in the engine of modern-day rankings. Having reviewed a large body of empirical studies on rankings, we have identified three dimensions along which rankings have evolved decisively once being produced by organizations: publication frequency; handling increasingly complex tasks; and audience engagement. We illustrated how, in contrast to rankings produced by individuals, organizations are better equipped to publish rankings on a continual basis, handle the increasingly complex production process, generate considerable degrees of attention by addressing larger and more diverse audiences, and develop mechanisms to respond to their criticism. In short, there is reason to believe that the organizational production of rankings provides an elementary and hitherto overlooked infrastructure undergirding the discursive institutionalization of rankings. On a theoretical level, therefore, our analysis suggests that we are well-advised to reconnect insights on the legitimacy of institutions as promoted by the "new" institutionalism with the "older" institutionalism's emphasis on the effectiveness and efficiency of organization in order to make sense of the pervasiveness of institutions such as rankings.

Acknowledgments

We wish to thank the thematic issue editors and three anonymous reviewers for their comments and suggestions on earlier versions of the article. All errors are our own.

Conflict of Interests

The authors declare no conflict of interests.

References

Ashforth, B., Sluss, D. M., & Harrison, S. H. (2007). Socialization in organizational contexts. *International Review of Industrial and Organizational Psychology*, *2007*, 1–70. https://doi.org/10.1002/9780470753378.ch1

Bahrami, R., & Meyer, M. (2019). Was motiviert die Rankingproduktion? Das Fallbeispiel der PIPA-Rankings [What motivates the production of rankings? The example of PIPA]. In L. Ringel & T. Werron (Eds.), *Rankings: Soziologische Fallstudien* [Rankings: Sociological case studies] (pp. 1–24). Wiesbaden: Springer VS.

Bandura, R. (2011). *Composite indicators and rankings: Inventory 2011 (UNDP/ODS Working Paper)*. New York, NY: United Nations Development Programme, Office of Development Studies.

Barnard, C. I. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press.

Brankovic, J., Ringel, L., & Werron, T. (2018). How rankings produce competition: The case of global university rankings. *Zeitschrift Für Soziologie*, *47*(4), 270–288. https://doi.org/10.1515/zfsoz-2018-0118

Brankovic, J., Ringel, L., & Werron, T. (2019). *Theoriz-*

ing university rankings: A comparative research perspective (Working Paper SFB 1288, No. 2). Bielefeld: Bielefeld University. Retrieved from https://pub.uni-bielefeld.de/record/2939616

Buckermann, P. (2019). *Die Vermessung der Kunstwelt: Quantifizierende Beobachtungen und plurale Ordnungen der Kunst* [Measuring the art world: Quantified observations and the plurality of orders in the arts]. Weilerswist-Metternich: Velbrück Wissenschaft.

Christin, A. (2018). Counting clicks: Quantification and variation in web journalism in the United States and France. *American Journal of Sociology*, *123*(5), 1382–1415. https://doi.org/10.1086/696137

Coase, R. H. (1937). The nature of the firm. *Economica*, *4*(16), 386–405. https://doi.org/10.1111/j.1468-0335.1937.tb00002.x

de Rijcke, S., Wallenburg, I., Wouters, P., & Bal, R. (2016). Comparing comparisons: On rankings and accounting in hospitals and universities. In J. Deville, M. Guggenheim, & Z. Hrdličková (Eds.), *Practising comparison: Logics, relations, collaborations* (pp. 251–280). Manchester: Mattering Press.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160.

Dörre, K., Lessenich, S., & Singe, I. (2013, April 27). German sociologists boycott academic ranking. *Global Dialogue: Magazine of the International Sociological Association*, *3*(3). Retrieved from http://globaldialogue.isa-sociology.org/german-sociologists-boycott-academic-ranking

Drori, G. S., Meyer, J. W., & Hwang, H. (Eds.). (2006). *Globalization and organization: World society and organizational change*. Oxford: Oxford University Press.

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, *113*(1), 1–40.

Espeland, W. N., & Sauder, M. (2016). *Engines of anxiety: Academic rankings, reputation, and accountability*. New York, NY: Russell Sage Foundation.

Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. *European Journal of Sociology/Archives Européennes de Sociologie*, *49*(3), 401–436. https://doi.org/10.1017/S0003975609000150

Esposito, E., & Stark, D. (2019). What's observed in a rating? Rankings as orientation in the face of uncertainty. *Theory, Culture & Society*, *36*(4), 3–26. https://doi.org/10.1177/0263276419826276

Fligstein, N., & McAdam, D. (2012). *A theory of fields*. New York, NY: Oxford University Press.

Freedom House. (2020). Freedom in the world 2020 methodology. *Freedom House*. Retrieved from https://freedomhouse.org/reports/freedom-world/freedom-world-research-methodology

Furnari, S. (2014). Interstitial spaces: Microinteraction settings and the genesis of new practices between institutional fields. *Academy of Management Review*, *39*(4), 439–462. https://doi.org/10.5465/amr.2012.0045

Global Integrity. (2009). Johann Lambsdorff retires the corruption perceptions ind gartnerex. *Global Integrity*. Retrieved from https://www.globalintegrity.org/2009/09/18/post-452

Gutterman, E. (2014). The legitimacy of transnational NGOs: Lessons from the experience of Transparency International in Germany and France. *Review of International Studies*, *40*(2), 391–418. https://doi.org/10.1017/S0260210513000363

Hammarfelt, B., de Rijcke, S., & Wouters, P. (2017). From eminent men to excellent universities: University rankings as calculative devices. *Minerva*, *55*(4), 391–411. https://doi.org/10.1007/s11024-017-9329-x

Jeacle, I., & Carter, C. (2011). In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. *Accounting, Organizations and Society*, *36*(4/5), 293–309. https://doi.org/10.1016/j.aos.2011.04.002

Jensen, C. B. (2011). Making lists, enlisting scientists: The bibliometric indicator, uncertainty and emergent agency. *Science & Technology Studies*, *24*(2), 64–84.

Jin, G. Z., & Whalley, A. (2007). *The power of attention: Do rankings affect the financial resources of public colleges?* (Working Paper No. 12941). https://doi.org/10.3386/w12941

Kelley, J., & Simmons, B. (2014). *The power of performance indicators: Rankings, ratings and reactivity in international relations* (APSA 2014 Annual Meeting Paper). Washington, DC: American Political Science Association. Retrieved from https://papers.ssrn.com/abstract=2451319

Kornberger, M., & Carter, C. (2010). Manufacturing competition: How accounting practices shape strategy making in cities. *Accounting, Auditing & Accountability Journal*, *23*(3), 325–349. https://doi.org/10.1108/09513571011034325

Kornberger, M., Pflueger, D., & Mouritsen, J. (2017). Evaluative infrastructures: Accounting for platform organization. *Accounting, Organizations and Society*, *60*, 79–95. https://doi.org/10.1016/j.aos.2017.05.002

Krause, M. (2018). How fields vary. *The British Journal of Sociology*, *69*(1), 3–22. https://doi.org/10.1111/1468-4446.12258

Kühl, S. (2015). The diffusion of organizations: The role of foreign aid. In B. Holzer, F. Kastner, & T. Werron (Eds.), *From globalization to world society: Neo-institutional and systems-theoretical perspectives* (pp. 258–278). New York, NY: Routledge.

Kunkel, B. W. (1915). The distinction of college graduates. *School and Society*, *2*(35), 316–324.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Lim, M. A. (2018). The building of weak expertise:

The work of global university rankers. *Higher Education*, *75*(3), 415–430. https://doi.org/10.1007/s10734-017-0147-8

Maanen, J. V., & Schein, E. H. (1979). Toward a theory of organizational socialization. *Research in Organizational Behavior*, *1*, 209–264.

Manning, P. K. (2008). Goffman on organizations. *Organization Studies*, *29*(5), 677–699. https://doi.org/10.1177/0170840608088767

Mau, S. (2019). *The metric society: On the quantification of the social* (1st ed.). Cambridge: Polity Press.

Mehrpouya, A., & Samiolo, R. (2016). Performance measurement in global governance: Ranking and the politics of variability. *Accounting, Organizations and Society*, *55*, 12–31. https://doi.org/10.1016/j.aos.2016.09.001

Meyer, J. W., & Bromley, P. (2013). The worldwide expansion of "organization." *Sociological Theory*, *31*(4), 366–389.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*(2), 340–363.

Münch, R. (2014). *Academic capitalism: Universities in the global struggle for excellence*. London: Routledge.

Orlikowski, W. J., & Scott, S. V. (2013). What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector. *Organization Science*, *25*(3), 868–891. https://doi.org/10.1287/orsc.2013.0877

Pollock, N., & D'Adderio, L. (2012). Give me a two-by-two matrix and I will create the market: Rankings, graphic visualisations and sociomateriality. *Accounting, Organizations and Society*, *37*(8), 565–586. https://doi.org/10.1016/j.aos.2012.06.004

Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Powell, W. W., & DiMaggio, P. J. (1991). *The new institutionalism in organizational analysis*. Chicago, IL: University of Chicago Press.

Rindova, V. P., Martins, L. L., Srinivas, S. B., & Chandler, D. (2018). The good, the bad, and the ugly of organizational rankings: A multidisciplinary review of the literature and directions for future research. *Journal of Management*, *44*(6), 2175—2208. https://doi.org/10.1177/0149206317741962

Ringel, L. (2018). Boundaries of visibility in the age of transparency: An integrative conceptualization. In L. Ringel, P. Hiller, & C. Zietsma (Eds.), *Toward permeable boundaries of organizations?* (pp. 55–79). Bingley: Emerald Publishing.

Ringel, L., & Werron, T. (in press). Where do rankings come from? A historical-sociological perspective on the history of modern rankings. In A. Epple, W. Erhart, & J. Grave (Eds.), *Practices of comparing: Ordering and changing the worlds*. Bielefeld: Bielefeld University Press.

Roose, K. D., & Andersen, C. J. (1970). *A rating of graduate programs*. Washington, DC: American Council on Education.

Schmidt-Burkhardt, A. (2005). *Stammbäume der Kunst, Zur Genealogie der Avantgarde* [Familiy trees of the arts: The genealogy of avantgarde]. Berlin and Boston, MA: De Gruyter.

Selznick, P. (1996). Institutionalism 'old' and 'new.' *Administrative Science Quarterly*, *41*(2), 270–277.

Simon, H. A. (1945). *Administrative behavior: A study of decision-making processes in administrative organization*. New York, NY: Free Press.

Simon, H. A. (1957). *Models of Man*. New York, NY: Wiley.

Spoerhase, C. (2018). Rankings: A pre-history. *New Left Review*, *114*, 99–113.

Stack, M. (2016). *Global university rankings and the mediatization of higher education*. London: Palgrave Macmillan.

Stinchcombe, A. L. (1997). On the virtues of the old institutionalism. *Annual Review of Sociology*, *23*(1), 1–18. https://doi.org/10.1146/annurev.soc.23.1.1

Stinchcombe, A. L. (2001). *When formality works: Authority and abstraction in law and organizations*. Chicago, IL: Chicago University Press. Retrieved from https://www.press.uchicago.edu/ucp/books/book/chicago/W/bo3622920.html

Strang, D., & Meyer, J. W. (1993). Institutional conditions for diffusion. *Theory and Society*, *22*(4), 487–511. https://doi.org/10.1007/BF00993595

Weber, M. (1947). *The theory of social and economic organization*. New York, NY: Oxford University Press. (Original work published ca. 1915)

Webster, D. S. (1986). *Academic quality rankings of American colleges and universities*. Springfield, IL: Charles C. Thomas.

Webster, D. S. (1992). Reputational rankings of colleges and universities, and individual disciplines and fields of study, from their beginnings to the present. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol 8, pp. 234–304). New York, NY: Agathon Press.

Werron, T. (2014). On public forms of competition. *Cultural Studies–Critical Methodologies*, *14*(1), 62–76. https://doi.org/10.1177/1532708613507891

Werron, T., & Ringel, L. (2017). Rankings in a comparative perspective: Conceptual remarks. In S. Lessenich (Ed.), *Geschlossene Gesellschaften: Verhandlungen des 38. Kongresses der Deutschen Gesellschaft für Soziologie* [Closed societies: Proceeding of the 38th Conference of the German Sociological Association] (pp. 1–10). Essen: DGS.

Wilbers, S. (2019). Grenzarbeit im Kunstbetrieb: Zur Institutionalisierung des Rankings Kunstkompass [Boundary work in the arts: The institutionalization of the ranking Kunstkompass]. In L. Ringel & T. Werron (Eds.), *Rankings: Soziologische Fallstudien* [Rankings: Sociological case studies] (pp. 57–87). Wiesbaden: Springer VS.

World Bank Group. (2018). *Doing business 2018: Reforming to create jobs—United States* (Working Paper No. 121076). Washington, DC: The World Bank. Retrieved from http://documents.worldbank.org/curated/en/252431510213684171/Doing-Business-2018-reforming-to-create-jobs-United-States

About the Authors

Leopold Ringel is a Lecturer at the Faculty of Sociology of Bielefeld University, Germany. His current research focuses on the unintended consequences of organizational transparency and quantification, rankings in particular. He has published in *Organization Studies* and *Research in the Sociology of Organizations*. He holds a PhD from Heinrich-Heine-University Duesseldorf, Germany.

Jelena Brankovic is a Postdoctoral Researcher at the Faculty of Sociology, Bielefeld University, Germany. Her current research focuses on the role rankings and other forms of comparisons play in the institutional dynamics within and across sectors, with a particular attention to the developments at the global level. She holds a PhD in Sociology from Ghent University, Belgium.

Tobias Werron is Professor of Sociological Theory at the Faculty of Sociology, Bielefeld University, Germany, and SFB 1288 Deputy Spokesperson since 2019. His research focuses on globalization, competition, rankings, and violent conflicts.

Politics and Governance (ISSN: 2183–2463)
2020, Volume 8, Issue 2, Pages 48–57
DOI: 10.17645/pag.v8i2.2584

Article

Keeping One's Shiny Mercedes in the Garage: Why Higher Education Quantification Never Really Took Off in Germany

Maarten Hillebrandt

Faculty of Law, University of Helsinki, 00014 Helsinki, Finland; E-Mail: maarten.hillebrandt@helsinki.fi

Submitted: 29 October 2019 | Accepted: 29 November 2019 | Published: 9 April 2020

Abstract
The cybernetic dream of regulatory 'dashboard control' has taken off in the German higher education system. Both government regulators and university managers are engaged in the creation of waves of increasingly fine-grained quantitative data. Yet a wide range of recent case studies of the German higher education sector attest that in spite of this 'datafication' frenzy, the impact of the collected data mass on regulatory and managerial decision-making capacities seems to have remained relatively limited. This article explores why, in spite of the considerable investment in quantitative data infrastructures in the German higher education sector, this did not result in significant overt analytical capacity building. It explores three hypotheses: 1) a legal hypothesis according to which quantification is curbed by legal protections under the *Rechtsstaat*; 2) a dysfunctionality hypothesis which holds that decision makers reject quantification as a flawed and impracticable pursuit; and 3) an egalitarian federalism hypothesis which argues that Germany's federal states seek to prevent commensurability to avoid comparison and competition. The article finds that, in spite of its inconspicuousness, quantification indeed does inform various central decision-making processes. However, different legal, political, and relational factors prompt decision makers to engage in a hybrid, tempered and, overall, untransparent application of numerical data.

Keywords
administrative capacity; data; education; Germany; quantification

Issue
This article is part of the issue "Quantifying Higher Education: Governing Universities and Academics by Numbers" edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

1. The Shiniest Car: Unused in the Garage?

The cybernetic dream of regulatory 'dashboard control' has taken off in higher education systems across the world (de Boer, Enders, & Schimank, 2008; Espeland & Sauder, 2016; Hood, James, Peters, & Scott, 2004, Chapter 3). Germany is no exception to this trend. Over the past three decades, the sector has engaged in the creation of waves of increasingly fine-grained data, which often takes on a numerical form (Franzen, 2018; Huber & Hillebrandt, 2019; Kleimann, 2016). Since about 10 years ago, university administrations, too, have begun to expand their performance data regarding teaching and research activities. The construction of information systems, quality assurance mechanisms, and core data sets is thus the talk of the town in German higher education (Biesenbender & Hornbostel, 2016; Seyfried & Pohlenz, 2017).

Surprisingly however, the impact of the collected data mass seems to have remained relatively limited in German higher education. Indeed, the initiation of numerical data collection projects with steering potential, followed by the relative disregard of that data in decision making processes appears to be a recurrent feature. To be sure, quantitative data are used for purposes of steering and reaching decisions on such issues as budgeting, research funding, hiring, and programme accreditation (Huber & Hillebrandt, 2019; Kleimann & Hückstädt, 2018; Leibner, 2017; Oberschelp, 2017). However, the impact of such quantification is largely displaced by

pre-existing arrangements. This even appears to be the case when such data are allegedly included in decisional procedures. Most German states' performance-based budgeting policy, for example, is based on only a few indicators, and capped to maximal year-on-year fluctuations, making the metric relatively inconsequential (Oberschelp, 2017, p. 109; Schimank, 2009, p. 134).

Some forms of quantification appear to have barely caught on. Higher education regulators and managers, for example, generally consider article-level metrics, which have mushroomed over the past decade, too context-specific and complex for decision-making frameworks (Franzen, 2018). Meanwhile, faculty appointment committees rather orient themselves on the intangible notion of research reputation (Kleimann & Hückstädt, 2018). (Re)accreditation procedures for study programmes, in turn, rely foremost on qualitative peer review (Schneijderberg & Steinhardt, 2018), and data collection activities to improve teaching quality are met with mistrust by university teachers (Seyfried & Pohlenz, 2017, p. 98). Finally, university rankings, produced by newspapers and specialised organisations, are almost wholly ignored by regulators and university managers alike (Leibner, 2017, p. 48).

The above brief overview, though impressionistic, offers a first indication that, even when universities and government regulators are quite interested in the establishment of number-based data infrastructures (Prenzel & Lange, 2017, p. 18), the impact of such data on decision making appears to be faltering. Decision making is decoupled from data evidence, objects appear difficult to quantify, or comparison between objects is considered problematic and is thus resisted. This article explores why the 'shiny Mercedes' of elaborate quantitative data in the German higher education sector has not resulted in analytical capacity building to the extent that it did elsewhere.

2. Connecting Quantification and Analytical Capacity

Quantification has for some time been associated with managerialism in the higher education sector and beyond. However, in the case of Germany, the link between quantification and analytical capacity, which is a precondition of managerialism, appears to be weak. This section develops a number of hypotheses as to why this might be the case.

2.1. Analytical Capacity as a Precondition of Managerialism

The past decades have seen the rise of the new public management in higher education sectors in Western Europe and beyond (de Boer et al., 2008). The new public management paradigm is widely perceived to be underpinned by a neoliberal programme for the public sector, in particular higher education, in which efficiency, accountability and a consumer orientation moved to the foreground as central guiding principles (Power, 1997, p. 43; Shore & Wright, 2000, p. 60). The diagnosis of neoliberal managerialism sparked a debate about the extent to which higher education sectors in different countries, not least in Germany, actually follow its central tenets (de Boer et al., 2008; Hood et al., 2004; Schimank, 2005). This article takes a step back from this discussion, to consider a *necessary precondition* for a managerialist outlook in higher education: the collection of and systematic engagement with performance and quality information for steering universities, or *analytical capacity* (Head, 2008, p. 2; Wollmann, 1989, p. 233). The process by which *numerical* indicators are used to effect impact on managerial or regulatory decision making (de Boer et al., 2008, pp. 37–40; Parrado, 2014, p. 88; Power, 1997, pp. 98–104), has been defined as *quantification* (Huber & Hillebrandt, 2019; for an alternative, more encompassing definition of quantification, see Espeland & Sauder, 2016, p. 21). The relation between quantification and analytical capacity may vary in intensity from one context to another. In some instances, the two may be tightly coupled to the extent that numerical indicators determine decisional outcomes. In these cases, we might say that the managerialist dream of 'governance by numbers' is fully realised (Huber & Hillebrandt, 2019, p. 260).

Various criteria however need to be fulfilled before we can speak of truly impactful analytical capacity (see Figure 1). Crucially, a *data infrastructure* must be in place. Data infrastructures typically require experts who collect and compile analytical data in a systematic manner (Parrado, 2014). Moreover, to speak of governance by numbers, that data needs to be predominantly *cast in quantitative terms*.

Quantitative data infrastructures can only enhance analytical capacity when they come about in a *clear decision-making context*. Governing by numbers goes beyond the mere observation and consideration of data. Instead, the indicators influence decisions in a direct way. The focus of analytically oriented decision makers is on the question: "What happens if we change these settings?" (Head, 2008, p. 1). As such, there must be at least a *plausible linkage between the collected data and the decisional output*. The more numerical representations alter decisional outcomes, the stronger the analytical capacity that quantification provides (Parrado, 2014, p. 89). In other words, "What matters is that moment when numbers oust judgment, or at least marginalise it or limit its operation to specific domains" (Kurunmäki, Mennicken, & Miller, 2016, p. 395). In practice however, decision makers' attitude towards data often appears more driven by circumstantial conditions than by the computing capabilities to which they have access as such (Head, 2008, p. 9; Parrado, 2014, p. 93).

Previous research found that while German federal government departments took active steps to enhance their data processing capacities, the connection between the resultant knowledge base and decision making remained weak (Wollmann, 1989, pp. 261–262). This study

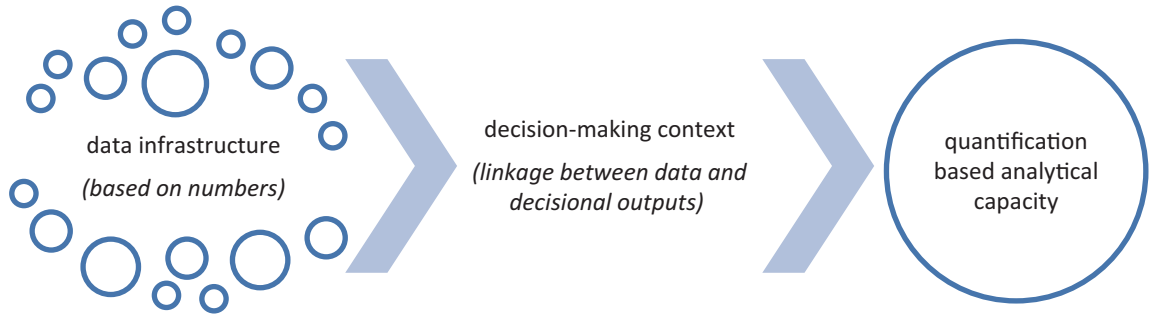

Figure 1. Governing by numbers: Quantification-based analytical capacity.

seeks to test and update these findings in the context of the considerably more digitised (Parrado, 2014, p. 86), decentralised (Schimank, 2005) and autonomised (Kleimann, 2016) German higher education sector.

2.2. Obstacles to Quantification-Based Analytical Capacity

The picture that emerges from recent literature sketches German higher education as a system that has, over the past years, produced a 'shiny new Mercedes' worth of new data infrastructures. However, the Mercedes appears mostly to stand in the garage. The collected data hardly make up a decisive factor in managerial and regulatory decision making. If German higher education decision makers have the numbers, why do they not let them govern? In order to explain this picture, this article considers three alternative hypotheses derived from different literatures.

According to the first explanation, here described as the *legal hypothesis*, the general desire among regulators and managers to rely on quantification is significantly curbed by existing legal arrangements, leading to a 'driving ban.' This view derives from the comparative policy literature, which describes the German decision-making context as a corporatist system built on a legalistic order, as opposed to the more liberal marketist Anglo-Saxon or social-democratic Scandinavian systems (Esping-Andersen, 2009). Although a direct comparison of the analytical capacity engendered by quantification in different higher education systems goes beyond the scope of this article, and the treatment of this hypothesis must therefore remain exploratory, it is here noted that this view is well-anchored in the higher education literature, which typically characterises German higher education as a system with strong academic rights entrenchment (Hood et al., 2004; Hüther & Krücken, 2013, p. 316; Seyfried & Pohlenz, 2017, pp. 99–100; see also de Boer et al., 2008, for further elaboration of this theme). This is believed to create a relatively resilient and reform-resistant governance system, undermining advances of governance by numbers.

A second (and alternative) explanation is the *dysfunctionality hypothesis*. According to this hypothesis, regulators and managers recognise the general value of quantitative data, but consider them unfit for incorporation in decision-making instruments. Negative experiences or general scepticism lead to the gradual side-tracking of indicators in the steering system, giving them a 'flat tire.' This hypothesis derives from a long-standing critical rationalist strand in the public management literature, which argues that management instruments—including data infrastructures—when unwieldy, must either be repaired or decoupled from decision making (Bouckaert & Balk, 1991; Van Thiel & Leeuw, 2002). Governing by numbers has been associated a host of dysfunctionalities associated with technical, definitional, and behavioural deficits (Oberschelp, 2017, pp. 112–118; Van Thiel & Leeuw, 2002, p. 270). The literature on (higher education) quantification has long recognised this problematic side to quantification-based steering, highlighting various potential unforeseen and perverse aspects in a variety of contexts (Espeland & Sauder, 2016; Huber & Hillebrandt, 2019; Strathern, 2000). It is less clear how German regulators and managers relate to such commonly reported observations.

The third hypothesis, that of *egalitarian federalism*, is rooted in the literature on federal politics (Lijphart, 1984/2010). This literature identifies a competitive dynamic both among the affiliated component states and between the decentral and the central government layers. It builds on the observation that German higher education makes up a devolved policy area (Dobbins & Knill, 2016, p. 72), and as such effectively functions as a governance quilt of sixteen independent and loosely cooperating systems (Hüther & Krücken, 2013, p. 308). Local differences in preference and political outlook lead to the absence of a coherent overarching steering logic, akin to a car with 'scattered parts.' The absence of a powerful coordinating centre (the federal government) means that decisions towards convergence take place on a lowest common denominator basis. It may be expected that this protective structure hinders efforts at coordinated quantification, as the latter opens new conflict lines regarding decisional autonomy.

The three hypotheses are shown in Table 1. Beyond these hypotheses, it also needs to be established whether previous scattered observations of the decoupling of analytical capacity from decision making can actually be generalised across the system. It is for exam-

Table 1. 'The shiny Mercedes': Why higher education quantification never really took off.

Hypothesis	Description	Situation
Legal hypothesis	'driving ban'	Influence of quantification is curbed by (constitutional) legal protections
Dysfunctionality hypothesis	'flat tire'	Failure of quantification leads to its gradual side-tracking
Egalitarian federalism hypothesis	'scattered parts'	Disconnect between state steering instruments, marginalising quantification

ple imaginable that the different hypotheses interact, or only apply under specific circumstances. For this reason, the hypotheses are treated in an exploratory manner, as scenarios that aid theorisation.

3. Research Approach

In order to investigate the perceived dynamics around different quantitative data infrastructures, the article considers the managerial and regulatory situation of universities in three German federal states, whereby the focus lies on decision makers and knowledge creators and brokers. Taken together, the selected states (whose identity is suppressed for reasons of respondent anonymity) are broadly representative of the available variety in German higher education sector along several variables, including number of public universities, expenditure (Deutschland in Zahlen, 2017), professor-student ratios (Destatis, 2018, p. 26), and political-ideological legacies (Lanzendorf & Pasternack, 2009; Schimank, 2009). Between them, the selected states enrol a sizeable proportion of all German students (Federal Statistical Office, 2018).

The analysis builds on data collected for a European comparative study on quantification in the public sector, namely the Open Research Area project entitled 'Quantification, Administrative Capacity and Democracy,' funded in Germany by the German Research Foundation (project number 627097). Part of the 'Quantification, Administrative Capacity and Democracy' project work consisted of the creation of a register of current quantification-based regulatory instruments and related policy documents, as well as interviews with various actors in the higher education sector. Regulatory instruments were inventoried on the basis of a review of relevant academic publications in German higher education journals, reports, policy documents, and websites produced by the central actors in the German higher education sector, including those produced by various authorities such as the science and education ministries of the selected states and the federal government, relevant state-affiliated agencies and institutes such as the statistics offices of the selected states and the federal government, the German Council of Science and Humanities, the German Research Foundation, the Accreditation Council, and the German Centre for Higher Education Research and Science Studies, and finally, the central administrations of the selected universities. The interviews in turn served to establish which quantification instruments are most salient from a regulatory and managerial perspective, and to certify that no instruments were overlooked. As quantitative data infrastructures presented by (commercial) third parties were found to have a negligible bearing on regulatory and managerial decision making, they are not treated in detail in the analysis (see however Ringel, Brankovic, & Werron, 2020; Krüger, 2020).

For this article, 15 interviews with 19 university managers and state regulators are included. The managers interviewed (12) occupy positions at central and faculty level in six universities (4 in state A, 1 each in states B and C) that vary in terms of size (in student numbers), age (founding year), wealth (in terms of budget per student and third-stream funding), and teaching pressure (students per professor). The selected regulators (7) occupy positions in the science and education ministries, or in directly related agencies. An overview of interviewees can be found at the end of this article. The analysis consisted of three steps. First, a baseline inventory of quantification-based data infrastructures at the university, state, and federal levels was conducted. Second, the background and functioning of these infrastructures were explored with interviewees, to identify to what extent and in what ways these infrastructures are coupled to decisional processes of steering and management. The third and final step consisted of establishing the presence or absence of the obstacles as formulated in the three hypotheses, both at university and state level. The following section will consider the findings more closely, whereby the focus lies on the third step of the analysis (identifying the presence/absence of evidence of the three hypotheses). References in this section remain limited to quotations (details on interview-based evidence can be obtained from the author).

The analysis also has a number of shortcomings. First, the single-country focus means that the study cannot verify whether identified explanations are specific to the German case, or whether they are more widely generalisable. Further (comparative) case studies would be required to corroborate this aspect. Second, due to space constraints, the reported analysis can only summarily engage with regulatory differences per level and pertaining to functional differentiation of activity (teaching/research). However, thorough efforts are made to

clarify the scope of empirical claims. Third, although efforts were made to include a wide variety of local conditions through a most-different cases and within-cases selection, the data derived from the interviews may not be fully representative of all public universities and states in the German higher education context.

4. Quantification-Based Analytical Capacity: The Invisible Mercedes

A baseline inventory of data infrastructures corroborates the starting position of this article that the German higher education sector is indeed undergoing a process of 'datafication' (Table 2). This development is nearly universally confirmed by the respondents.

At the same time, many of the same respondents argue that direct governance by numbers remains a marginal phenomenon, suggesting that data infrastructures are indeed decoupled from decision making. Respondents identify performance-based budgeting as the major exception to this trend. However, the financial impact and administrative burden of performance-based budgeting are everywhere limited (see also Section 4.2). The downside of explicitly decoupling data analyses from regulation is that indicators remain inconsequential and obtain a largely rhetorical character:

> It is always the best for all, if it continues the way it is. Good, you could say, then we also don't need to collect new numbers, then we actually don't need to do anything. But I'll just say, the influence of the media on politics is getting stronger. It is true that a science minister today gets asked more and more about various subjects, and then of course he wants to have

Table 2. The 'datafication' of the German higher education sector.

Locus	Type of data infrastructure	
	Teaching	Research
Faculties	Indicator dashboards (levels of advancement differ strongly)*	
		Performance contracts with faculties or professors (some quantification)
University administrations	Central data analysis departments with business intelligence systems, financial health data and data warehouses**	
		Performance contracts with faculties or professors (some quantification)
State ministries		Performance contracts with universities (some quantification)
	Performance-based university budgeting system relying on quantitative indicators	
	Data visualisations to enhance transparency regarding performance	
	Instruments for charting universities' student absorption capacities and average operational teaching costs per discipline area	
Federal Science Council		'Research rating': distribution data on different indicators
		'Core Data Set': promoting standardisation of various definitions for different reporting purposes***
State and federal statistics offices	Various mandatorily reportable data per state, discipline or semester	
(Semi-) public research bodies	Graduate surveys at national and state level (partially quantified)	

Notes: * This partially depends on the extent to which discipline areas are already accustomed to working with quantitative data; ** I.e., data concerning research projects (e.g., external funding attracted and data from the German Research Foundation Funding Atlas), concerning publications (journal impact and citations, often divided by the number of professors), and data concerning teaching modules (including students attracted, number of credits handed out, student satisfaction); *** Most universities are currently aligning their internal data accordingly.

> something to answer. (Respondent 2, translation by the author)

In summary, the inventory of quantification instruments confirms that number-based data infrastructures are indeed present, growing, and actively promoted in key areas of German higher education governance, but are hardly coupled directly to decisional outcomes. Decision makers tend to predominantly use number-based data infrastructures as a way of 'getting grip' on a managerial context. What holds them from taking the final leap, by letting the numbers decide for them? The following analysis suggest that the answer to this question lies in a number of specific legal, administrative, and federal political conditioning factors.

4.1. Legal Rights and Obligations: Acting in the Shadow of Legalistic Culture

According to the legal hypothesis, quantification-based analytical capacity is incongruent with legal protections guaranteed by the German *Rechtsstaat*. Respondents, both university managers and government regulators, frequently asserted that they would like to instigate data-driven decision making, but are legally prevented from doing so. The force of legal constraints is underlined by a series of significant (constitutional) court interventions over the past decades setting out the rules of engagement in higher education. In one state, for example, the competitive element between universities was removed from the performance-based budgeting system after criticism from its Court of Auditors. Another legal constraint on the use of quantitative data for strategic purposes is data protection law, which in most cases prohibits the dissemination of individually attributable data:

> We can't even say how high the study discontinuation rate is….Because we cannot verify study careers since, for reasons of data protection, such data are not collected. (Respondent 2, translation by the author)

Clearly, a strict interpretation of the notion of data protection sets limits on the possibility of equity-enhancing quantification going in the direction of affirmative action to ensure *de facto* enrolment and graduation of ethnic minorities, akin to policies developed in other countries, notably England. There, an Office for Fair Access keeps explicit track of the diversity of student populations within specific universities for regulatory purposes (United Kingdom Department of Education and Skills, 2004).

The constitutional protection of academic autonomy is also strong. In particular, tenured professors enjoy extensive protections from interference. In practice, this makes it difficult to motivate them to provide performance data. A weaker variety of the autonomy argument applies to the university organisation as a whole. In some cases, universities refuse to deliver requested data, because they are unconvinced about the knowledge creators' methodology. In regulatory relations where they cannot avoid quantification, universities will lobby hard for the adoption of the least consequential form. One area in which they have been unsuccessful, is that of calculating capacity for creating study places. Here, court action to protect citizens' legal right to higher education has created a *numerus clausus* system building on an elaborate formula for absorption capacity. In other choices, for example human resources, university management is constrained by detailed regulation protecting professional prerogatives.

To a large extent however, reluctance to rely on overt quantification to arrive at managerial or regulatory decisions is related to *semi*-legal argumentation. When driving rules are heavily regulated, it may seem more opportune to leave the car inside altogether. In many cases managers or regulators would be free to engage in data-driven decision making, they seek to avoid confrontations, which in turn strengthens their interlocutors' negotiating position. No law, for example, prevents state ministries from drawing a far more rigid connection between quantified performance and university budgets. However, consensualism requires the downplaying of performance transparency to avoid embarrassment for 'losing' universities. University managers could propose sharper internal financial distribution formulas. However, their position of 'leader among peers' means that university managers generally steer clear of openly confrontational quantification:

> I think, one should always bear in mind the specificity that we are not in the open economy, where you can harshly say: "You have not reached your target, tomorrow you can find yourself another place to work." I am very happy…that I am analysing it externally. (Respondent 9, translation by the author)

The legal hypothesis is thus far from a generalisable feature. The law is rather invoked by academics or universities to highlight specific prerogatives, than to resist quantification per se. However, when too conspicuous, quantification is deemed to intervene undesirably in the existing system of interdependencies of the higher education system. The reliance on data infrastructure is thus accorded a relatively invisible place in the background.

4.2. Dysfunctional Data: Inducing Caution, but Not Rejection

The dysfunctionality hypothesis presents an alternative explanation: Decision makers' believe that data-driven decision making does not work. Indeed, regulators are well aware of the negative side effects of governing by numbers, and there is strong reluctance to couple quantitatively measured performance to significant budgetary decisions:

> As a university, I can't from one day to the next lose €10 million….Most of the money is locked into wages and salaries. In existing contracts. They have to be honoured as well. And [the buildings] need to be heated, and the lights have to be on, and then there's hardly anything left. Then you can cut a few research assistants and…some project budget, but that's not all too much. (Respondent 2, translation by the author)

This perception does not threaten the perpetuation of the performance-based budget model, which functions unobtrusively in the background (see Section 4, introduction). The already limited impact of the model is further constrained by additional demand-based teaching quality and pay-per-student funding schemes introduced to address the pressing financial situation in the sector. Both schemes are largely input-oriented without regard for output.

Quantification-based analytical capacity entails various difficulties related to technical aspects. Required data may be unreliable or unavailable in a standardised, longitudinal form. In some cases, the desired data are altogether absent:

> The coalition agreement of the [current state] government…stated that the drop-out rate should be reduced by 20 percent….We read that the next day, and said: "How should that work? We don't even know what the *current* drop-out rate is." (*laughs*) (Respondent 3, translation by the author)

A further technical aspect is the feasibility of data infrastructure building. For effective steering, decision makers require up-to-date numbers. However, in many cases data collection is a resource-intensive and time-consuming affair. Beyond technical aspects, there is the definitional question what indicators are required for particular policy objectives. Available data may not fit goals or be irrelevant for some academic disciplines. For this reason, most budgetary allocation models, whether imposed on or within universities, avoid complex and controversial data. Decision makers struggle with this problem and arrive at different solutions in different states:

> We always found this discussion difficult. I observed that in the other states as well: with research, they are always quick [with operationalisations]….That's not controversial. In the area of teaching, there is a discussion in our state as well as in others: what is actually the right indicator? The question is, when I count graduates, do I actually capture anything that represents teaching quality? (Respondent 14, translation by the author)

In terms of behavioural aspects of quantification, it is widely believed that too much transparency (e.g., through ranking) leads to inequitable outcomes and must remain limited. This stands in stark contrast to experiences in many countries, including the United States, where private-party rankings hold strong sway over the sector (Espeland & Sauder, 2016). While such rankings also exist in Germany, decision makers largely refuse to make prestige and resource allocation dependent on them. At the same time, both regulators and managers point out that risks can be hedged successfully, for example by capping maximal budgetary fluctuations, or by incorporating discretionary elements. Particularly university managers see quantification as a hybrid form amenable to experimentation (see also Huber, 2020). As such, a distinction can be made between data-driven and data-*informed* decision making. While the former makes decisions explicitly dependent on particular data pictures, be it through the setting of thresholds, minimum values, classifications, or more formulaic, even algorithmic, numerical composites, the latter merely uses such pictures as an inspiration that is expressly non-committal and open-ended.

In sum, the findings do not convincingly support the dysfunctionality hypothesis. Decision makers show an awareness of the potential dysfunction of quantification-based analytical capacity, but this does not stop them from using numbers to inform their decision making. Aware of the risks involved, they strive for practical feasibility, modesty of expectations, and hybridity. Even with a flat tire, the Mercedes might get you further than going by foot, albeit perhaps rather via backroads than the highway. This promotes an untransparent and somewhat detached incorporation of key indicators into traditional decision-making processes.

4.3. Resistance to National Commensuration: Transparency Limits Engagement

The egalitarian federalism hypothesis focusses particularly on the attitude of state regulators, and applies only to German higher education inasmuch as it is regarded as a single system. The broad support for different national data initiatives shows that the states are willing to cooperate actively to make standardisation and numerical comparison more feasible, albeit reluctantly. The research rating was mainly thought of as a counterweight to international rankings that were questioned on grounds of methodology and their fit with the German higher education landscape. Indeed, for a variety of reasons, Germany consistently undershoots its expected target based on sector size and budget in these rankings.

Still, coordinated national 'datafication' efforts remain modest. One regulator in a more forward state speculated that the kind of performance visualisation tool that existed in his state could never be mainstreamed nationally:

> We are after all a large state, we have a whole row of universities here that can also be compared….We would very much like to do this for the entire federal republic, but we happen to have this federalism

> and every state knows best in this area, so in that sense…we're not going to get that. (Respondent 3, translation by the author)

Some evidence suggests that when national quantification efforts create too much visibility, this diminishes support. When the pilot study of the research rating exercise performed in a limited number of discipline areas disappeared 'in the drawer,' observers speculated this might have been due to the undesirable transparency that this created about states' university performance. While there was not much to be gained from good performance, widespread under-performance would most likely trigger negative media attention and sour relations between individual states and their universities. Such political sensitivities play out differently in other federal states, such as the United Kingdom, where higher education policies are devolved, but where the vast majority of universities are based in England, where the introduction of various transparency-enhancing regulatory instruments in a top–down fashion has been going on for over two decades now.

One result of the limited enthusiasm for direct state comparison is the tendency of states to disown national quantification processes, by emphasising universities' role as stakeholders in the process:

> In the context of the pilot phase, the states did not refer the findings back to themselves….Even to the contrary, if I remember correctly, in the [institutional] process, they even retrospectively said: "We as states actually have no opinion at all about whether we need the research rating or not. We'll hear what the universities have to say about it." (Respondent 8, translation by the author)

Beyond state efforts, comparison of state higher education sectors also emerges from other knowledge brokers. The Funding Atlas published by the German Research Foundation, however focuses throughout on universities and disciplines. This is illustrated by the fact that, seemingly deliberately, not a single table breaks findings down by state (German Research Foundation, 2018). This differs from the Federal Statistics Office, which in its regular publications offers state comparisons on a wide range of indicators, from on-time graduations to the academic staff-to-student ratio or expenditure per student and professor (Destatis, 2018), or the German Rectors' Conference, which compiles data on e.g., the diversity of study programmes on offer and *numerus clausus* courses per state (German Rectors' Conference, 2018).

The findings suggest that while the states endorse diplomacy-like cooperation towards opening up the German university landscape in numbers and indicators, they withdraw when things get too transparent. Beyond federal programmes, the contribution of knowledge creators and brokers such as university platforms, the German Research Foundation, and the Federal Statistics Office to the creation of a quantified competitive field remains limited and removed from state or federal decision-making processes. Perhaps the German states prefer to regard higher education quantification as a Mercedes garage workshop: harmless when left to tooling fans, potentially harmful when the cars are actually taken out for a drive.

5. Conclusion

Regulators and managers in the German higher education system have a 'shiny Mercedes' worth of number-based data infrastructures at their disposal, but rarely couple this data to decision making in the form of quantification-based analytical capacity. In 2009, the German sociologist Schimank stated that "policy learning implies a willingness to learn, [which] is, whether on the professors' side or the state's side, still hardly a given" (Schimank, 2009, p. 136, translation by the author). Ten years later, this picture seems to have shifted, at least as regards state regulators and university managers. Both groups show a broad interest in 'learning from the numbers,' and letting the new insights inform their decision making. However, they do so in a manner that remains tied into institutionalised traditional decision-making arrangements. As a result, quantification-based analytical capacity functions in a largely tempered, hybrid, and untransparent manner that has little to do with algorithmic understandings of direct 'governance by numbers.'

The three different hypotheses discussed in the analysis further clarify the manner in which legal arrangements, administrative prudence, and federal politics constrain certain manifestations of quantification-based decision making. While legal arrangements and the German *Rechtsstaat* squarely protect prerogatives such as academic and institutional autonomy, this in itself does not prohibit regulatory or managerial decision making through quantification. For example, university management typically enters professorial appointment or salary negotiations with a quantified overview of the person's performance and achievements. That said, in many decisional contexts a degree of administrative prudence and consensualism prevents regulators and managers from engaging in forms of quantification that are considered too disruptive, for example by creating undesired transparency or antagonising particular actors. The German states, for example, endorse data-informed cooperation only to the extent that quality differences are not shown too transparently, to avoid stirring undesired competition.

In the limited number of cases where steering is directly indicator-based, decision makers shun data complexity, generally prioritising simplicity, reliability and feasibility over multidimensionality, completeness and precision. A prime example are performance-based budgeting systems, which function on the basis of only a few indicators and whose potential effects are capped in advance. The role played by data infrastructures in in-

forming decision making in the background is harder to capture. Particularly managers inform themselves with a wide variety of data that differ significantly from one university to the next.

It thus emerges that data infrastructures are in fact present in a variety of central decision-making settings. Yet, where quantification is most consequential, it is also most watered down by discussion forums and/or expert evaluations. As such it functions to a large extent under the radar, in the sense that the strength of its influence is not easily demonstrable. To outer appearances, the 'shiny Mercedes' of analytical capacity appears to stay mostly in the garage. In reality, regulators and managers across Germany might take it out for an inconspicuous drive more often than has previously been thought. Further research may help uncover the structural consequences of the growth of quantification in German higher education governance, and the extent to which it has encroached on traditional academic self-governance, while national comparative research could demonstrate whether the German Mercedes experience is unique, or rather similar to that of the Rolls-Royces and Lamborghinis of other higher education systems.

Acknowledgments

This article was first presented at the workshop 'Quantifying Higher Education: Origins, Production, Consequences,' held at Bielefeld University on 28–29 March 2019, where it benefitted from elaborate constructive feedback from the participants. In addition, the author would like to acknowledge valuable comments provided by Michael Huber, Luis Sanz Menendez, Laura Cruz Castro, Tobias Werron, and the anonymous reviewers, as well as Fabio Schiebel for his assistance with collecting and processing the data.

Conflict of Interests

The author declares no conflict of interests.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Biesenbender, S., & Hornbostel, S. (2016). The research core dataset for the German science system: Developing standards for an integrated management of research information. *Scientometrics*, *108*(1), 401–412.

Bouckaert, G., & Balk, W. (1991). Public productivity measurement: Diseases and cures. *Public Productivity and Management Review*, *15*(2), 229–235.

de Boer, H., Enders, J., & Schimank, U. (2008). Comparing higher education governance systems in four European countries. In N. C. Soguel & P. Jaccard (Eds.), *Governance and performance of education systems* (pp. 35–54). Wiesbaden: Springer.

Destatis. (2018). Hochschulen auf einem Blick [Universities in overview]. *Destatis*. Retrieved from https://www.destatis.de/Migration/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/BroschuereHochschulenBlick.html

Deutschland in Zahlen. (2017). Staatliche Hochschulausgaben je Studierenden: In Euro [State university expenditure per student: In Euros]. *Deutschland in Zahlen*. Retrieved from www.deutschlandinzahlen.de/tab/bundeslaender/bildung/bildungsausgaben/staatliche-hochschulausgaben-je-studierenden

Dobbins, M., & Knill, C. (2017). Higher education governance in France, Germany, and Italy: Change and variation in the impact of transnational soft governance. *Politics and Society*, *36*(1), 67–88.

Espeland, W. N., & Sauder, M. (2016). *Engines of anxiety: Academic rankings, reputation, and accountability*. New York, NY: Sage.

Esping-Andersen, G. (2009). *The three worlds of welfare capitalism*. Cambridge: Polity Press.

Federal Statistical Office. (2018). Studierende Wintersemester 2017/2018 [Enrolled students, winter semester 2017/2018]. *Federal Statistical Office*. Retrieved from www.statistikportal.de/de/studierende

Franzen, M. (2018). Die digitale Transformation der Wissenschaft [The digital transformation of science]. *Beiträge zur Hochschulforschung*, *40*(4), 8–28.

German Rectors' Conference. (2018). Statistische Daten zu Studienangeboten an Hochschulen in Deutschland [Statistical data regarding programme offer at universities in Germany]. *German Rectors' Conference*. Retrieved from https://www.hrk.de/fileadmin/redaktion/hrk/02-Dokumente/02-03-Studium/02-03-01-Studium-Studienreform/HRK_Statistik_BA_MA_UEbrige_WiSe_2019_20_finale_internet.pdf

German Research Foundation. (2018). *Förderatlas 2018: Kennzahlen zur öffentlich Finanzierten Forschung in Deutschland* [Research Atlas 2018: Indicators of publically funded research in Germany]. Bonn: German Research Foundation. Retrieved from https://www.dfg.de/sites/foerderatlas2018/download/dfg_foerderatlas_2018.pdf

Head, B. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration*, *67*(1), 1–11.

Hood, C., James, O., Peters, B. G., & Scott, C. (Eds.). (2004). *Controlling modern government: Variety, commonality and change*. Cheltenham: Edward Elgar Publishing.

Huber, M. (2020). Steered by numbers: How quantification differentiates the reform of a German university. *Politics and Governance*, *8*(2), 26–35.

Huber, M., & Hillebrandt, M. (2019). "Pay for promise" in higher education: The influence of NPM on resource allocation in German universities. *Historical Social Research*, *44*(2), 247–269.

Hüther, O., & Krücken, G. (2013). Hierarchy and power: A

conceptual analysis with particular reference to new public management reforms in German universities. *European Journal of Higher Education*, *3*(4), 307–323.

Kleimann, B. (2016). *Universitätsorganisation und präsidiale Leitung: Führungspraktiken in einer multiplen Hybridenorganisation* [University organisation and rectoral management: Leadership practices in a multi-hybridities organisation]. Wiesbaden: Springer.

Kleimann, B., & Hückstädt, M. (2018). Auswahlkriterien in Berufungsverfahren: Universitäten und Fachhochschulen im Vergleich [Selection criteria in appointment procedures: Universities and politechnic colleges compared]. *Beiträge zur Hochschulforschung*, *40*(2), 20–46.

Krüger, A. K. (2020). Quantification 2.0? Bibliometric infrastructures in academic evaluation. *Politics and Governance*, *8*(2), 58–67.

Kurunmäki, L., Mennicken, A., & Miller, P. (2016). Quantifying, economising, and marketising: Democratising the social sphere? *Sociologie du Travail*, *58*, 390–402.

Lanzendorf, U., & Pasternack, P. (2009). Hochschulpolitik im Ländervergleich [A state comparison of university policies]. In J. Bogumil & R. G. Heinze (Eds.), *Neue Steuerung von Hochschulen: Eine Zwishenbilanz* [The new steering of universities: Taking stock] (pp. 13–28). Berlin: Edition Sigma.

Leibner, T. (2017). University governance and rankings: The ambivalent role of rankings for autonomous accountability and competition. *Beiträge zur Hochschulforschung*, *39*(3/4), 30–51.

Lijphart, A. (2010). *Patterns of democracy*. New Haven, CT: Yale University Press. (Original work published 1984)

Oberschelp, A. (2017). Das Fächerrating des Landes Hessen: Ausgestaltung, Einsatz und Leistungsfähigkeit eines Instruments im Rahmen der Landeshochschulsteuerung [The disciplines rating in the state of Hessen: Design, implementation and performance of an instrument in the context of state university regulation]. *Beiträge zur Hochschulforschung*, *39*(1), 104–129.

Parrado, S. (2014). Analytical capacity. In M. Lodge & K. Wegrich (Eds.), *The problem-solving capacity of the modern state: Governance challenges and administrative capacities* (pp. 86–104). Oxford: Oxford University Press.

Power, M. (1997). *The audit society: Rituals of verification*. Oxford: Oxford University Press.

Prenzel, M., & Lange, S. (2017). Evidenzbasierte Governance von Organisationen in Forschung und Lehre: Erwartungen an die Wissenschafts und Hochschulforschung [Evidence-based governance of organisations in research and teaching: Expectations of science studies and higher education governance]. *Beiträge zur Hochschulforschung*, *39*(1), 10–20.

Ringel, L., Brankovic, J., & Werron, T. (2020). The organizational engine of rankings: Connecting "new" and "old" institutionalism. *Politics and Governance*, *8*(2), 36–47.

Schimank, U. (2005). 'New public management' and the academic profession: Reflections on the German situation. *Minerva*, *43*, 361–376.

Schimank, U. (2009). Governance-Reformen nationaler Hochschulsysteme: Deutschland in internationaler Perspektive [Governance reforms of national higher education systems: Germany in international perspective]. In J. Bogumil & R. G. Heinze (Eds.), *Neue Steuerung von Hochshulen: Eine Zwishenbilanz* [The new steering of universities: Taking stock] (pp. 123–137). Berlin: Edition Sigma.

Schneijderberg, C., & Steinhardt, I. (2018). Accreditation of X qualities instead of quality X: A normative analysis of criteria of the German higher education accreditation regime. *Higher Education Policy*, *32*(1), 5–28.

Seyfried, M., & Pohlenz, P. (2017). Zwischen Wunsch und Wirklichkeit: Qualitätsmanagement als weiches Diziplinierungsinstrument [Between wish and reality: Quality management as soft disciplining instrument]. *Beiträge zur Hochschulforschung*, *39*(3/4), 96–115.

Shore, C., & Wright, S. (2000). Coercive accountability: The rise of audit cultures in higher education. In M. Strathern (Ed.), *Audit cultures: Anthropological studies in accountability, ethics, and the academy* (pp. 57–89). London: Routledge.

Strathern, M. (Ed.). (2000). *Audit cultures: Anthropological studies in accountability, ethics, and the academy*. London: Routledge.

United Kingdom Department of Education and Skills. (2004). Guidance letter from Secretary of State for Education and Skills. *United Kingdom Department of Education and Skills*. Retrieved from www.offa.org.uk/about/background

Van Thiel, S., & Leeuw, F. (2002). The performance paradox in the public sector. *Public Performance and Management Review*, *25*(3), 267–281.

Wollmann, H. (1989). Policy analysis in West Germany's federal government: A case of unfinished governmental and administrative modernization? *Governance*, *2*(3), 233–266.

About the Author

Maarten Hillebrandt is Postdoctoral Researcher at the Erik Castrén Institute, University of Helsinki, where he is involved in a research project on the transparency of EU decision making. Previously, he participated in the international comparative ORA-funded research project 'Quantification, Administrative Capacity and Democracy' (2016–2019), in which researchers from the London School of Economics, Paris MinesTech, Bielefeld University, Helmut Schmidt University Hamburg, and Leiden University were co-collaborators. Maarten specialises in EU institutional law and politics, government transparency policy, and quantification.

Politics and Governance (ISSN: 2183–2463)
2020, Volume 8, Issue 2, Pages 58–67
DOI: 10.17645/pag.v8i2.2575

Article

Quantification 2.0? Bibliometric Infrastructures in Academic Evaluation

Anne K. Krüger [1,2]

[1] Berlin Brandenburg Academy of Sciences and Humanities, 10117 Berlin, Germany; E-Mail: anne.krueger@bbaw.de
[2] Department of Science Studies, Humboldt University Berlin, 10099 Berlin, Germany

Submitted: 22 October 2019 | Accepted: 3 February 2020 | Published: 9 April 2020

Abstract
Due to developments recently termed as 'audit,' 'evaluation,' or 'metric society,' universities have become subject to ratings and rankings and researchers are evaluated according to standardized quantitative indicators such as their publication output and their personal citation scores. Yet, this development is not only based on the rise of new public management and ideas on 'the return on public or private investment.' It has also profited from ongoing technological developments. Due to a massive increase in digital publishing corresponding with the growing availability of related data bibliometric infrastructures for evaluating science are continuously becoming more differentiated and elaborate. They allow for new ways of using bibliometric data through various easily applicable tools. Furthermore, they also produce new quantities of data due to new possibilities in following the digital traces of scientific publications. In this article, I discuss this development as quantification 2.0. The rise of digital infrastructures for publishing, indexing, and managing scientific publications has not only made bibliometric data become a valuable source for performance assessment. It has triggered an unprecedented growth in bibliometric data production turning freely accessible data about scientific work into edited databases and producing competition for its users. The production of bibliometric data has thus become decoupled from their application. Bibliometric data have turned into a self-serving end while their providers are constantly seeking for new tools to make use of them.

Keywords
bibliometrics; big data; digital infrastructures; higher education; quantification

Issue
This article is part of the issue "Quantifying Higher Education: Governing Universities and Academics by Numbers" edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

1. Introduction

Current observations—discussed as the 'audit' (Power, 1999), 'evaluation' (Dahler-Larsen, 2012), or 'metric society' (Mau, 2019)—indicate a lack of societal trust in the performance of public organizations and the individuals working in them. Power describes this "audit explosion" as a need for more control through constant performance measurement that is based on "a certain set of attitudes or cultural commitments to problem solving" (Power, 1999, p. 4) which have been transferred from private companies to public organizations—such as universities. With the rise of new public management and economic calculations about the return on public (see Schimank, 2005) or—in the case of the US (see Espeland & Sauder, 2016)—private investment into science and higher education, universities have become subject to internal evaluations (see Hillebrandt, 2020; Huber, 2020; Matthies & Simon, 2008) and external ratings and rankings (see Brankovic, Ringel, & Werron, 2018; Hazelkorn, 2011; Espeland & Sauder, 2016) according to their performance in research and increasingly also in teaching (see Times Higher Education, 2018). Consequently, this has also led to an increasing evaluation of the performance of individual researchers building on standardized quantitative indicators such as their publication output and their personal citation scores (see de Rijcke, Wouters, Rushforth, Franssen, & Hammarfelt, 2016; Waltman, van Eck, Visser, & Wouters, 2016).

Yet, this development towards performance measurement as problem solving is not only based on a shift in political attitudes and societal orders of worth but has

profited from ongoing technological developments that provide new ways of producing and assessing data about performances. The aim of the article therefore is to highlight the significant changes in data production and assessment that can be witnessed, in particular, in the context of bibliometric research evaluation. Already in 1964, Eugene Garfield started the Science Citation Index that was based on automated data processing (with handwritten punch cards) and the use of the first IBM computers (see Wouters, 1999, pp. 26–27). Due to a massive increase in digital publishing and the growing availability of publication metadata such as author names, reference lists, author affiliations, or funding organizations, such bibliometric infrastructures are increasingly becoming more differentiated and elaborate. They allow for new ways of using these data through various easily applicable tools. Furthermore, bibliometric infrastructures also produce new quantities of data due to new possibilities in following the digital traces scientific publications leave when people within or outside of academia engage with them, for instance, by viewing an article or downloading it from a journal website. The counts of such metadata about article usage are discussed as alternative metrics or 'altmetrics' (see Franzen, 2015; Haustein, Sugimoto, & Larivière, 2015).

In this article, I suggest to address these new developments from a critical data perspective (see boyd & Crawford, 2012) to highlight how new ways of data production affect questions of data application and usage. First, I provide some insights into the development of bibliometrics from research information to research evaluation. I highlight how the "competition for expertise" (de Rijcke & Rushforth, 2015, p. 1955) on research evaluation has brought new providers of bibliometric infrastructures onto the stage. Second, I discuss the increasing possibilities of bibliometric infrastructures from a sociomaterial perspective (see Orlikowski & Scott, 2008). I argue that the assemblage of infrastructures, their providers, and their users influences how science and research practice are understood and thus how 'research performance' is measured. And third, drawing on insights from critical data studies, I suggest discussing these developments as quantification 2.0. I argue that in applying these infrastructures, it has not only become a standardized practice to turn qualitative characteristics into quantitative metrics by making different things such as the individual work of researchers commensurable (see Espeland & Stevens, 2008). Moreover, I highlight that the rise of digital infrastructures for publishing, indexing, and managing scientific publications has triggered an unprecedented growth in bibliometric data production turning freely accessible data about scientific work into edited databases. Bibliometric data production has thereby become decoupled from questions of utility and usability. Instead, bibliometric data has turned into a self-serving end while their providers are constantly seeking for new tools to make use of them.

2. Quantifying Science as 'Research Performance'

In 2008, Wendy Espeland and Mitchell Stevens called for a "sociology of quantification" because they witnessed "the spread of quantification" defined as "the production and communication of numbers" and "the significance of new regimes of measurement" (Espeland & Stevens, 2008, p. 402; for a detailed overview on the different lines of study on quantification, see Mennicken & Espeland, 2019). University ratings and rankings either for distributing public money or for informing prospective students have played a significant role in transforming "qualities into quantities" and "difference into magnitude" by reducing and simplifying "disparate information into numbers that can be easily compared" (Espeland & Stevens, 1998, p. 316). One of the commonly used indicators of such rankings is the publication output and the respective citation score (see Hazelkorn, 2011, pp. 32–37; Taubert, 2013). Even in the Times Higher Education Europe Teaching Ranking, which was released for the first time in 2018, "papers-to-staff ratio" adds to the overall ranking result of the teaching quality of a university by 7.5 percent (Times Higher Education, 2018, p. 29).

Yet, first ideas behind the collection of metadata of journal articles such as author names and references were not about research evaluation but about research information (for an encompassing history of bibliometrics, see Wouters, 1999). As a reaction towards the profound increase of scientific publications (see Wouters, 1999, pp. 59–60), in the 1960s, Eugene Garfield started to build up the Science Citation Index with a public grant funded by the US–American National Institutes of Health. The goal behind the Science Citation Index was to facilitate the search for publications in medicine and natural sciences among researchers. It was thus designed as a tool for enabling researchers to collect information about latest work for keeping track with new developments.

When information scientists like Derek de Solla Price (1963) but also Eugene Garfield himself (1979) recognized that bibliometrics provided possibilities for studying science as such they began to develop this field of research further. Based on the Science Citation Index, they started to explore the development of the science system and within particular disciplinary fields. As Taubert describes it:

> Not the single citation but instead the emerging patterns from the analysis of masses of citations were of interest that allowed for insights into the importance of particular research groups, institutions, national research systems or the dynamic of research fields. (Taubert, 2013, p. 183, translation by the authors)

The search for patterns furthermore allowed identifying the journals that were deemed most relevant in a particular research field (see Garfield, 2007). This information was summarized in the *Journal Citation Report* which was

designed for helping librarians in choosing the journals that scientists cite the most.

Initially designed as a tool for researchers and librarians and for research about science, bibliometrics furthermore began to feature more and more prominently in science policy. Bibliometric calculations became a relevant 'judgment device' (Karpik, 2010) for evaluating research organizations and individual researchers and, consequentially, for allocating resources within the science system. De Rijcke and Rushforth argue that:

> [The bibliometric] field had managed to create a demand for their measures—not only by supplying on-demand data and data-handling techniques but also by making sure their products were promoted as policy relevant information that decision makers could use strategically. (2015, p. 1955)

Publication output and citation scores thus turned into indicators for measuring research performance with different intensity according to the national and/or regional research system.

However, since the 1990s, concerns have constantly grown among the bibliometrics community about how these data are used calling for 'responsible metrics' (see, for an overview, Aksnes, Langfeldt, & Wouters, 2019; Ràfols, 2019; Wilsdon, 2015; Wilsdon et al., 2017). The Leiden Manifesto for research metrics (see Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015) is but one of the prominent examples (see also San Francisco Declaration on Research Assessment, 2013, or the Hong Kong Manifesto for Assessing Researchers, 2019) of how the bibliometrics community struggles with the problem that bibliometric indicators for research evaluation are no longer only used by experts from the bibliometrics community but often by research organizations and policy institutions themselves (see Hammarfelt & Rushforth, 2017; Leydesdorff, Wouters, & Bornmann, 2016).

3. The 'Competition for Expertise'

Despite critical debates within the bibliometrics community, "an increasingly intense globalized competition for expertise" (de Rijcke & Rushforth, 2015, p. 1955) on research evaluation has grown. The provision of bibliometric data for evaluation purposes has turned into "a crowded marketplace" (de Rijcke & Rushforth, 2015, p. 1956). The Institute of Scientific Information, which was founded by Garfield in 1960, is no longer the only provider of bibliometric data. Currently, there is an increasing number of databases that provide bibliometric information because the conditions for data production have changed. While Garfield and his colleagues still had to select and process data on journals, articles, authors and references manually (see Wouters, 1999), bibliometric data production has become facilitated through digital publishing and automated data collection. Due to these developments new databases have developed. There are open initiatives such as Crossref, which seeks to provide information on links between publications, funders, preprints or datasets across different publishers. Furthermore, there are community-based subject-specific databases such as Astrophysics Data System or MathSciNet that operate as bibliographic databases for the search of relevant literature, but also include citation information. Yet, there is also a growing field of databases that is provided by commercial providers. Besides Web of Science, which was initiated by Garfield and is currently owned by Clarivate Analytics, commercial databases such as Scopus and Dimensions have emerged. They also belong to major companies, namely Elsevier and Digital Science. These companies do not only provide bibliometric databases, but furthermore own a large infrastructure of tools to generate information from these data about people, organizations or even entire countries regarding their research activities. However, there are significant differences between these databases.

Web of Science is a collection of databases that has emerged from the Science Citation Index and the Social Science Citation Index that were launched in 1964 and 1965 by the Institute of Scientific Information. From 1975 on, the Arts and Humanities Citation Index was added. Since 1990, Web of Science also contains the Conference Proceedings Citation Index and since 2005 it enlists also books in the Book Citation Index. Only recently, from 2015 onwards, Web of Science has also integrated the Emerging Sources Citation Index to "make content important to funders, key opinion leaders, and evaluators visible…even if it has not yet demonstrated citation impact on an international audience" (Clarivate Analytics, 2017a). Web of Science therefore comprises a plurality of datasets thereby attempting to capture different disciplines with their distinct styles of publishing research as journal articles, books, or conference proceedings.

Scopus was launched by the publisher Elsevier in 2004. In contrast to Web of Science, which indexes only a specific set of supposedly key journals in their field, Scopus attempts to index the largest number of peer-reviewed literature possible. Yet, only publications that have a Digital Object Identifier are included. Elsevier also claims to control the selected journals for their quality through the Content Selection & Advisory Board that is made up of "an international group of scientists, researchers and librarians who represent the major scientific disciplines" which "is comprised of 17 Subject Chairs, each representing a specific subject field" (Elsevier, 2020). Scopus consists of only one database that, however, contains information on items that range from journals, books, and conference proceedings to patents and trade publications from all common disciplines.

The Holtzbrinck Publishing Group, who holds the majority of shares of Springer Nature and thus owns—similar to Elsevier—more than 2,500 English language journals, has recently become a new player in the bibliometric field. They own the company Digital Science.

Digital Science describes itself as "a technology company serving the needs of scientific and research communities at key points along the full cycle of research" (Bode, Herzog, Hook, & McGrath, 2018). In 2018, Digital Science launched Dimensions as another bibliometric database that seeks to cover a broader range of sources than Web of Science or even Scopus. Dimensions is described as "transcend[ing] existing tools and databases" by "bringing together…grants, publications, clinical trials and patents, consistently linked and contextualized" (Bode et al., 2018, p. 9). Dimensions is thus becoming a new prominent player in bibliometric data production and assessment.

Yet, due to technological developments in data collection, other players beyond academia and academic publishers have entered the bibliometric scene. Sweeping the entire academic web from digitally available journals to books and even websites in all languages and all countries, Google uses its search engine not only for supporting the search for academic literature (see Martín-Martín, Orduña-Malea, Ayllón, & López-Cózar, 2015). With GoogleScholar, which was set up in 2004, Google is furthermore able to extract and analyze the mentions of publications and its authors from all these sources. Disciplines like the social sciences and arts and humanities where coverage in the established databases like Web of Science or Scopus is still fragmentary and incomplete are thus discussed to have become more visible through it (see Bornmann, Thor, Marx, & Schier, 2016; Harzing & Alakangas, 2016). GoogleScholar is furthermore freely accessible. It has thereby become a convenient source for researchers and administrators to search for research but also for information on research performance.

The recent growth in the provision of bibliometric data thus highlights that there is already an ongoing competition for producing and providing data on academic publications from journal articles to conference reports as well as their analysis. Yet, while the coverage of these databases is continuously getting broader due to digitalization, quality issues arise that are prominently discussed in the bibliometrics community. The most prominent problem is that the providers of these databases give only very limited insights into how they collect and edit their data because this kind of information is their actual business secret. Dimensions and GoogleScholar claim to capture as much data as possible. This holds particular problems for the validity and reliability of their data. GoogleScholar, for instance, can only draw on publications that are available online and in contrast to Web of Science or Scopus it is impossible to acquire access to its database to get insights into which kinds of data are actually included. Prins, Costas, van Leeuwen, and Wouters (2016, p. 267) therefore highlight that "restrictions to the use of GoogleScholar are the intensive manual data handling and cleaning, necessary for a feasible and proper data collection" (see also Mingers & Meyer, 2017).

Web of Science and Scopus are instead more selective (see Stahlschmidt, Stephen, & Hinze, 2019, Chapter 3). Besides scientific quality, they also apply more formal criteria for selecting their data sources. Peer review and "ethical publication practices" as well as bibliographic information in English language are important selection criteria (Clarivate Analytics, 2017b; Elsevier, n.d.-a; Testa, n.d.). Yet, these selection criteria still leave some room for interpretation. How decisions are actually made on choosing journals and indexing them within which discipline remains opaque. In the case of journals categorized in the Emerging Sources Citation Index in Web of Science, they can pass "an initial editorial evaluation and can continue to be considered for inclusion in products such as the Science Citation Index Expanded, the Social Science Citation Index, and the Arts and Humanities Citation Index, which have rigorous evaluation processes and selection criteria" (Clarivate Analytics, 2017a). What this evaluation looks like, however, is not openly discussed. For Scopus, Taşkin, Doğan, Akça, Şencan, and Akbulut have shown that journals indexed in this database in some cases fail to have information e.g., about their publication ethics, malpractice management, or editorial policies publicly available although this information comprises part of Scopus' defined selection criteria (2015, as cited in Stahlschmidt et al., 2019). Journals might also be excluded from Scopus based on decisions made by the Scopus Content Selection and Advisory Board. These decisions on indexing journals have a significant impact on the results these databases produce. Despite attempts to integrate more journals and other publication formats these databases are "still less accurate for the social sciences and humanities than for other fields, and for certain regions such as Africa, Oceania and Central and South America" (Stahlschmidt et al., 2019, p. 10) due to an overrepresentation of English language publications and a predominant focus on journal publications (see also Mongeon & Paul-Hus, 2016).

4. Bibliometric Databases as Digital Infrastructures

In the bibliometrics community, these issues are mainly discussed as a methodological problem in terms of data quality and the construction of indicators. Yet, de Rijcke and Rushforth (2015) highlight that there is also an 'implementation problem' as the use of these data for evaluation purposes has already spread widely. Besides providing information for researchers, librarians, administrators, policy makers and funders, Elsevier also promotes Scopus as being used by influential ranking organizations such as Times Higher Education for their World University Rankings or the Shanghai Ranking Consultancy for the Best Chinese University Ranking Report (see Elsevier, n.d.-b). Clarivate Analytics supports the British Research Excellence Framework, Dimensions was recently used for the first time by the Nature Index 2019 annual tables (see Digital Science 2019), and GoogleScholar can already be used without any further restrictions by anybody interested in his or her personal metrics or the metrics of fellow researchers.

The bibliometrics community is therefore not the first reference any more for doing research evaluation as the providers themselves offer evaluations or evaluation tools (see also Jappe, Pithan, & Heinze, 2018; Petersohn & Heinze, 2018) that enable even 'lay persons' such as research managers and policy makers to do evaluations on their own. In particular, Clarivate Analytics, Elsevier, and Digital Science have built an encompassing digital infrastructure that produces and collects different sorts of data and metadata and processes and assesses them: Tools such as literature management systems like Menderley or EndNote facilitate the production of bibliometric data that do not only help to organize research literature but also provide information about the use of publications as well as formally correct and thus easily collectable citations. Furthermore, so called 'research intelligence solutions' are offered and promoted as enabling research managers from research institutions up to the policy level to gain, analyze, and also visualize information on the development of current research trends as well as on the performance of individual researchers, research groups, or research organizations. In addition, these tools are also designed to enable research managers to collect data about their own research organizations and to analyze and manage it.

Methodological problems in terms of data collection and assessment are still and have become even more a predominant issue as the use of such bibliometric infrastructures constantly spreads. Moreover, by collecting and managing data and facilitating the assessment of 'research performance,' these infrastructures play a performative role in creating what they seek to count and calculate, namely, a particular understanding of science and research practice. Hence, such digital infrastructures are never neutral but embody already in their design and the calculative models behind them a particular understanding about the world and about its users (see Krüger, Heßelmann, & Hartstein, in press; Mühlhoff, 2018). Already in 1999, Bowker and Star (1999, p. 230) have emphasized that "values, policies, and modes of practice become embedded in large information systems." In a similar vein, Winner has argued that:

> Machines, structures, and systems of modern material culture can be accurately judged not only for their contributions to efficiency and productivity, not merely for their positive and negative environmental side effects, but also for the ways in which they can embody specific forms of power and authority. (Winner, 1980, p. 121)

Building on these insights, Bowker, Elyachar, Mennicken, Miller, and Randa Nucho (2019, p. 1) highlight three important aspects of what they call "thinking infrastructures" defined as social or material infrastructures that "structure attention, shape decision-making and guide cognition." They are "valuation regimes that constitute orders of worth" (Bowker et al., 2019, p. 4) through definitions of success and failure; they make objects and practices "visible and available…for possible interventions" (Bowker et al., 2019, p. 4) thereby "establishing a distinct conception of the objects and objectives" of governance (p. 5). Digital infrastructures thus influence the practices they are supposed to support or reflect. They are performative because "they change the very nature of what it is to do work, and what work will count as legitimate" (Bowker & Star, 1999, p. 239).

In their research agenda on the role of sociomateriality in organization research, Orlikowski and Scott (2008) furthermore argue to go beyond the dichotomy of infrastructures and its users. Instead, they highlight that "people and things only exist in relation to each other" (Orlikowski & Scott, 2008, p. 455). Following Callon (1986) and Latour (1987) with their idea on actor-networks and the notion of 'performativity,' they claim that "entities (whether humans or technologies) have no inherent properties, but acquire form, attributes, and capabilities through their interpenetration" (Orlikowski & Scott, 2008, p. 455). The authors therefore think of the social and the material as "inherently inseparable" (p. 456) because they conjointly enact what this sociomaterial assemblage is about. Pollock, Williams, and Procter (2003) have applied this theoretical lens to a study on the construction and implementation of enterprise resource planning systems at universities. They demonstrate how this infrastructure is developed through its providers according to the perceived needs of the users, while the universities simultaneously adapt to this infrastructure, its integrated standardized processes and inscribed ideas about working practices. In this regard, infrastructure and universities mutually constitute each other.

Applied to the case of bibliometric infrastructures, this theoretical perspective allows us to see, first, that the assemblage of infrastructures, their providers and users does not produce stable constructs. Instead, bibliometrics are performed either as research information, research on research, or as research evaluation. Second, focusing on the performativity of these assemblages enables us to analyze how an understanding of science and research practice as "research performance" is constructed while attempting to measure and describe it.

In their comparative study of Web of Science and Scopus, Stahlschmidt et al. (2019) were able to demonstrate how different bibliometric databases construct a particular understanding of science and research practice leading to differences in their results. Building on a sample of German publications indexed in both databases, they found a "database-specific valuation of these publications" (Stahlschmidt et al., 2019, p. 64). Publications from the economic sector or from research institutes with a focus on applied sciences got better citation scores in Scopus than in Web of Science, while, conversely, in Web of Science organizations scored better that had a focus on basic research. Stahlschmidt et al. (2019) explain these differences through the impact that the respective content of each database has on the

valuation of a specific publication. In valuating a publication, Web of Science and Scopus draw on this content for "relating a publication to a specific environment of similar publications. Due to differences in coverage, Web of Science and Scopus apply different environments to appraise the same publication" (Stahlschmidt et al., 2019, p. 64). They therefore find that:

> Any differences in the valuation of the same content result from differences in the respective environment, i.e., the exclusive content. Hence a comparison of the diverging valuation of the same content does not inform on the content itself, but on the exclusive content causing any differences and therefore the databases themselves. (Stahlschmidt et al., 2019, p. 65)

Evaluation results thus depend on the assumptions about research that are inscribed in the databases as well as how users apply which database for which kind of evaluation purpose. Bibliometric infrastructures therefore have a particular impact on the production of knowledge about science as well as on scientific knowledge production as such. The assemblage of bibliometric infrastructures, their providers, and their users performs a particular understanding about the way research should be practiced and therefore effects which kind of research is consequently regarded as valuable.

5. Quantification 2.0

The growing abilities of bibliometric infrastructures due to digital publishing, however, do not only pose a problem in terms of performativity that influences data production and assessment. The digitalization of academic publishing has changed bibliometric data production as such. Data does not have to be produced manually anymore, but can be collected, processed, and furthermore assessed automatically. This has led to an increase in quantitative data production as such.

So far, research on evaluation has highlighted quantification in terms of "the production and communication of numbers" (Espeland & Stevens, 2008, p. 402; see also Heintz, 2010) that turn qualitative characteristics into quantitative metrics by making different things commensurable (see Espeland & Stevens, 1998). However, focusing on data production in bibliometrics shows that the production of data does not only follow from operationalizing qualitative differences in research performance in terms of quantitative output. Moreover, the competition for expertise highlights that these digital infrastructures contribute to what I suggest to call quantification 2.0: the decoupling of data production and data application.

In the literature on critical data studies phenomena such as 'big data' and 'datafication' and the automated use of data termed 'algorithmization' are already widely discussed. boyd and Crawford (2012, p. 663) define 'big data' as "a cultural, technological, and scholarly phenomenon" that results from "maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets" and "to identify patterns in order to make economic, social, technical, and legal claims." Similarly, Amoore and Piotukh (2015, p. 345) highlight that "the rise of big data witnesses a transformation in what can be collected or sampled as data, and how it can be rendered analyzable." Yet, most importantly, boyd and Crawford (2012) claim that 'big data' rests on some kind of "mythology," i.e., "the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy." The accumulation and automated analysis of large amounts of data has thus attained high credibility in the provision and objectivation of "digitally recorded, machine processable, easily agglomerated, and highly mobile" (Sadowski, 2019, p. 4) information about the social world. Fourcade and Healy (2017, p. 9) therefore claim that "modern organizations follow an institutional data imperative to collect as much data as possible." They argue that "it does not matter that the amounts collected may vastly exceed a firm's imaginative reach or analytic grasp" (Fourcade & Healy, 2017). Data is thus, as Sadowski puts it, "very often collected without specific uses in mind" (Sadowski, 2019, p. 4). Instead, "the assumption is that it will eventually be useful, i.e., valuable" (Fourcade & Healy, 2017, p. 13). Thus, metaphors such as 'data mining' or data as the 'new oil' have already spread widely understanding data as a new form of capital. Sadowski highlights that "the imperative…is to constantly collect and circulate data by producing commodities that create more data and building infrastructure to manage data" (Sadowski, 2019, p. 4). Producing, collecting, and processing data has become an "intrinsic motivation" (p. 4).

In the case of bibliometric databases, such developments display in the massive growth of digital publishing and the technological advancements in bibliometric infrastructures allow for collecting, processing, and assessing large amounts of data through automated processes (see de Rijcke & Rushforth, 2015; Taubert, 2013). Citations can now be collected and analyzed much easier through specific tools. Processes of standardization, e.g., of author names or organizations through the introduction of 'persistent identifiers,' simplifies data collection because they allow for the exact attribution of publications to authors and research organizations. Usage-based metrics such as, e.g., article views and downloads from journal websites, can be counted and analyzed as giving insights into the reception of publications beyond citations (see Haustein, Bowman, & Costas, 2016). Bibliometric tools, furthermore, produce data themselves that can be collected and analyzed for evaluation purposes. Taubert refers, for instance, to online reference managers that produce metadata about the usage of publications from storing and bookmarking to annotations by the reader (see Taubert, 2013, p. 25).

Providers of bibliometric infrastructures such as Clarivate Analytics, Elsevier, or Digital Science thus constantly seek to collect bigger amounts of data and metadata about publications with better data quality and to propose new ways of using and applying it. Companies such as Altmetric that is related to Digital Science or Plum Analytics, which belongs to Elsevier, have even started to collect data about the usage of scientific publications extending to mentions on Twitter and Facebook, among others (see Franzen, 2015). The valuation of bibliometric data furthermore displays in monetary transactions. When, in 1992, Thomson Reuters bought Web of Science from the Institute of Scientific Information, they paid $210 million (see Jayapradeep & Jose, 2017). When they resold the product to Clarivate Analytics in 2016, they received $3.55 billion for it (see Thomson Reuters, 2016). In addition, as already shown, the rise of new players such as Scopus, GoogleScholar, or recently Dimensions demonstrates that data on academic publications have become a valuable commodity. They capitalize on these data as they turn freely accessible and massively available data into edited databases. The problem of digital infrastructures is thus not only a question of data quality and of predefined assumptions inscribed in calculative models for data assessment. Moreover, the production of data has become a self-serving end as a profitable, privately owned and secretly kept business model while their providers are constantly seeking for new tools to make use of it.

6. Conclusion and Outlook

The possibilities of data production and assessment through bibliometric infrastructures have massively increased through digital publishing and automated data processing. In this article, I have argued that this does not only contribute to a quantification and evaluation of science and research practice as 'research performance.' Moreover, the rise of new providers and constantly more databases and related tools suggests what I have called quantification 2.0: the decoupling of data production and data application. Data production has become a self-serving end that is generating the development of new tools in search for a purpose. Political attitudes that understand evaluations as solving the problem of societal distrust in the performance of public organizations supported by increasingly complex digital infrastructures for data production and assessment have not only triggered a need for quantifying qualitative performances. Moreover, triggered through these developments, increasingly more quantifiable data are produced and processed which allow new ways to propose research evaluation to potential users.

From this, however, follows that even though or right because tools for bibliometric data production and assessment have become a business model for their providers, they have to address the needs of potential customers belonging to the science system. Their providers thereby profit from what boyd and Crawford have called "mythology" in terms of the perceived objectivity of big data and automated data analysis (boyd & Crawford, 2012, p. 663). Yet, the functionalities of newly designed bibliometric infrastructures and their characteristics still need to be explained and justified in a way to attract potential customers who are situated in the science system either as researchers and research managers or as policy makers, publishers, librarians or funding bodies. The providers of bibliometric infrastructures therefore have to propagate an understanding of science and research practice that resonates with the perception of their users.

In their research on "the extended practice of global software development," Campagnolo, Pollock, and Williams (2015) address the question how software development works when providers seek to enter new contexts with their products. They describe how software providers try to make sense of matters and customers that they do not yet know much about. They refer to the phenomenological concept of 'appresentation' as a way of taking "account of the needs of future customers and also of their current users of whom they have no direct knowledge" (Campagnolo et al., 2015, p. 150). The software providers thus design their products according to the imagined needs of potential customers they seek to address with it. In their study, Campagnolo et al. (2015) demonstrate how these appresentations affect the interactions between providers and customers in their search for a mutually shared understanding of the functionalities that the software needs to provide. In addition, they are able to show how the appresentation of the providers heavily influences how customers finally understand their own needs.

Regarding bibliometric infrastructures, the providers also have to work with an appresentation of anticipated needs of their customers when bringing new tools for data assessment on the market. These appresentations are already inscribed in the functionalities of bibliometric tools. Bibliometric infrastructures thus already display a particular understanding of science and research practice by enabling specific modes of observation and the presentation of respective results. Yet, this means simultaneously that the functionalities of bibliometric infrastructures also have to account for the ways in that their users understand 'research performance.' The appresentation of customer needs and how customers are supposed to be attracted therefore yields insights into the understandings of science and of research practice that are already inscribed in these bibliometric infrastructures and have a performative effect on the perception of their users.

It therefore seems promising for future research not only to keep track with recent technological developments, but, furthermore, to ask for the understandings of science and research practice that are displayed in the sociomaterial assemblage of infrastructures, their providers, and their users. We therefore need more re-

search, on the one hand, on the understandings of science and research practice that are inscribed in these infrastructures. On the other hand, we need more research about the application of these tools to understand how the mythology of 'better data, better decisions' affects how research is practiced and valued.

Acknowledgments

The author owes special thanks to Stephan Stahlschmidt, Sabrina Petersohn, Judith Hartstein, and Felicitas Heßelmann from the German Centre for Higher Education Research and Science Studies (DZHW) for critical debates and fruitful comments. The author also wants to thank the Academic Editors and the anonymous reviewers for their helpful comments on earlier versions of this article.

Conflict of Interests

The author declares no conflict of interests.

References

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, *9*(1). https://doi.org/10.1177/2158244019829575

Amoore, L., & Piotukh, V. (2015). Life beyond big data: Governing with little analytics. *Economy and Society*, *44*(3), 341–366.

Bode, C., Herzog, C., Hook, D., & McGrath, R. (2018). *Dimensions report: A guide to the dimensions data approach*. London: Digital Science and Research Solutions Inc. Retrieved from https://figshare.com/articles/A_Guide_to_the_Dimensions_Data_Approach/5783094

Bornmann, L., Thor, A., Marx, W., & Schier, H. (2016). The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute. *Journal of the Association for Information Science and Technology*, *67*(11), 2778–2789.

Bowker, G. C., Elyachar, J., Mennicken, A., Miller, P., & Randa Nucho, J. (2019). Introduction to thinking infrastructures. In M. Kornberger, G. Bowker, J. Elyachar, A. Mennicken, P. Miller, J. Randa Nucho, & N. Pollock (Eds.), *Thinking infrastructures* (pp. 1–13). Bingley: Emerald Group Publishing.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679.

Brankovic, J., Ringel, L., & Werron, T. (2018). How rankings produce competition: The case of global university rankings. *Zeitschrift für Soziologie*, *47*(4), 270–288.

Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St. Brieuc Bay. In J. Law (Ed.), *Power, action, and belief: A new sociology of knowledge?* (pp. 196–233). London and Boston, MA: Routledge & Kegan Paul.

Campagnolo, G. M., Pollock, N., & Williams, R. (2015). Technology as we do not know it: The extended practice of global software development. *Information and Organization*, *25*(3), 150–159.

Clarivate Analytics. (2017a). *Web of Science Core Collection: Emerging Sources Citation Index*. Philadelphia, PA: Clarivate Analytics. Retrieved from http://wokinfo.com/media/pdf/ESCI_Fact_Sheet.pdf

Clarivate Analytics. (2017b). Evaluation criteria for Web of Science Core Collection journals. *Clarivate Analytics*. Retrieved from https://clarivate.com/essays/evaluation-criteria-web-science-core-collection-journals

Dahler-Larsen, P. (2012). *The evaluation society*. Stanford, CA: Stanford University Press.

de Rijcke, S., & Rushforth, A. (2015). To intervene or not to intervene; is that the question? On the role of scientometrics in research evaluation. *Journal of the Association for Information Science and Technology*, *66*(9), 1954–1958.

de Rijcke, S., Wouters, P., Rushforth, A., Franssen, T., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use: A literature review. *Research Evaluation*, *25*(2), 161–169.

de Solla Price, D. (1963). *Little science, big science*. New York, NY: Columbia University Press.

Digital Science. (2019, June 19). Dimensions tracks normalized ranking for the first time in the Nature Index annual tables. *Digital Science*. Retrieved from https://www.digital-science.com/press-releases/dimensions-tracks-normalized-ranking-for-the-first-time-in-the-nature-index-annual-tables

Elsevier. (n.d.-a). Content policy and selection. *Elsevier*. Retrieved from https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection

Elsevier. (n.d.-b). Ranking organizations. *Elsevier*. Retrieved from https://www.elsevier.com/solutions/scopus/who-uses-scopus/ranking-organizations

Elsevier. (2020). *Scopus content coverage guide*. Amsterdam: Elsevier. Retrieved from https://www.elsevier.com/__data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf

Espeland, W., & Sauder, M. (2016). *Engines of anxiety: Academic rankings, reputation, and accountability*. New York, NY: Russell Sage Foundation.

Espeland, W., & Stevens, M. (1998). Commensuration as a social process. *Annual Review of Sociology*, *24*(1), 313–343.

Espeland, W., & Stevens, M. (2008). A sociology of

quantification. *Archives Europeennes de Sociologie*, *3*, 401–436.

Fourcade, M., & Healy, K. (2017). Seeing like a market. *Socio-Economic Review*, *15*(1), 9–29.

Franzen, M. (2015). Der Impact Faktor war gestern: Altmetrics und die Zukunft der Wissenschaft [The impact factor is old-fashioned: Altmetrics andd the future of science]. *Soziale Welt*, *66*(2), 225–242.

Garfield, E. (1979). *Citation indexing: Its theory and application in science, technology and humanities*. New York, NY: John Wiley & Sons.

Garfield, E. (2007). The evolution of the Science Citation Index. *International Microbiology*, *10*, 65–69.

Hammarfelt, B., & Rushforth, A. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, *26*(3), 169–180.

Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787–804.

Haustein, S., Bowman, T., & Costas, R. (2016). Interpreting "altmetrics": Viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication*: *A festschrift in honor of Blaise Cronin* (pp. 372–406). Berlin and Boston, MA: De Gruyter Mouton.

Haustein, S., Sugimoto, C., & Larivière, V. (2015). Guest editorial: Social media in scholarly communication. *Aslib Journal of Information Management*, *67*(3). https://doi.org/10.1108/AJIM-03-2015-0047

Hazelkorn, E. (2011). *Rankings and the reshaping of higher education: The battle for world-class excellence*. Houndmills: Palgrave Macmillan.

Heintz, B. (2010). Numerische Differenz: Überlegungen zu einer Soziologie des (quantitativen) Vergleichs [Numerical difference: Ideas on a sociology of (quantitative) comparison]. *Zeitschrift für Soziologie*, *39*(3), 162–181.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431.

Hillebrandt, M. (2020). Keeping one's shiny Mercedes in the garage: Why higher education quantification never really took off in Germany. *Politics and Governance*, *8*(2), 48–57.

Huber, M. (2020). Steered by numbers: How quantification differentiates the reform of a German university. *Politics and Governance*, *8*(2), 26–35.

Jappe, A., Pithan, D., & Heinze, T. (2018). Does bibliometric research confer legitimacy to research assessment practice? A sociological study of reputational control, 1972–2016. *PLoS ONE*, *13*(6). https://doi.org/10.1371/journal.pone.0199031

Jayapradeep, M., & Jose, B. (2017). Eugene Garfield: The ultimate impact factor. *Informatics Studies*, *4*(2), 7–24.

Karpik, L. (2010). *Valuing the unique: The economics of singularities*. Princeton, NJ: Princeton University Press.

Krüger, A. K., Heßelmann, F., & Hartstein, J. (in press). Bewertung in und durch digitale Infrastrukturen [Evaluation in and through digital infrastructures]. In T. Peetz & F. Meier (Ed.), *Bewertung und Organisation* [Evaluation and organization]. Wiesbaden: VS Springer.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Leydesdorff, L., Wouters, P., & Bornmann, L. (2016). Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use of indicators-a state-of-the-art report. *Scientometrics*, *109*(3), 2129–2150.

Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & López-Cózar, E. (2015). *Does Google Scholar contain all highly cited documents (1950–2013)?* Retrieved from https://arxiv.org/abs/1410.8464

Matthies, H., & Simon, D. (Ed.). (2008). *Wissenschaft unter Beobachtung* [Science under observation]. Wiesbaden: VS Verlag für Sozialwissenschaften.

Mau, S. (2019). *The metric society: On the quantification of the social*. Cambridge: Polity Press.

Mennicken, A., & Espeland, W. (2019). What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, *45*(1), 223–245.

Mingers, J., & Meyer, M. (2017). Normalizing Google Scholar data for use in research evaluation. *Scientometrics*, *112*(2), 1111–1121.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228.

Mühlhoff, R. (2018). Digitale Entmündigung und User Experience Design: Wie digitale Geräte uns nudgen, tracken und zur Unwissenheit erziehen [Digital disenfranchisement: How digital devices nudge and track us and how they teach us to become ignorant]. *Leviathan*, *46*(4), 551–574.

Orlikowski, W., & Scott, S. (2008). Sociomateriality: Challenging the separation of technology, work and organization. *The Academy of Management Annals*, *2*(1), 433–474.

Petersohn, S., & Heinze, T. (2018). Professionalization of bibliometric research assessment: Insights from the history of the Leiden Centre for Science and Technology Studies (CWTS). *Science and Public Policy*, *45*(4), 565–578.

Pollock, N., Williams, R., & Procter, R. (2003). Fitting standard software packages to non-standard organizations: The 'biography' of an enterprise-wide system. *Technology Analysis & Strategic Management*, *15*(3), 317–332.

Power, M. (1999). *The audit society: Rituals of verification*. Oxford and New York, NY: Oxford University Press.

Prins, A., Costas, R., van Leeuwen, T., & Wouters, P. (2016). Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. *Research Evaluation*, *25*(3), 264–270.

Ràfols, I. (2019). S&T indicators in the wild: Contextualization and participation for responsible metrics. *Research Evaluation*, *28*(1), 7–22.

Sadowski, J. (2019). When data is capital. Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1). https://doi.org/10.1177/2053951718820549

San Francisco Declaration on Research Assessment. (2013). *San Francisco Declaration on Research Assessment*. Retrieved from http://www.ascb.org/dora

Schimank, U. (2005). 'New Public Management' and the academic profession: Reflections on the German situation. *Minerva*, *43*(4), 361–376.

Stahlschmidt, S., Stephen, D., & Hinze, S. (2019). Performance and structures of the German science system. In Expertenkommission Forschung und Innovation (EFI) (Eds.), *Studien zum deutschen* Innovationssystem [Studies on the German Innovation System]. Berlin: Expertenkommission Forschung und Innovation (EFI).

Taşkin, Z., Doğan, G., Akça, S., Şencan, I., & Akbulut, M. (2015). *Does Scopus put its own journal selection criteria into practice?* Paper presented at the 15th International Conference of the International Society for Scientometrics and Informetrics, Istanbul, Turkey.

Taubert, N. (2013). Bibliometrie als Forschungsevaluation: Zur Konstitution und Funktionslogik wechselseitiger Beobachtung zwischen Wissenschaft und Politik [Bibliometrics as research evaluation: The constitution and logic of mutual observation between science and politics]. In J.-H. Passoth & J. Wehner (Eds.), *Quoten, Kurven und Profile: Zur Vermessung der sozialen Welt* [Quotas, curves and profiles: Measuring the social world] (pp. 179–204). Wiesbaden: Springer.

Testa, J. (n.d.). Journal selection process. *Clarivate Analytics*. Retrieved from https://clarivate.com/essays/journal-selection-process

Thomson Reuters. (2016). *Thomson Reuters announces definitive agreement to sell its intellectual property and science business to Onex and Baring Asia for $3.55 billion* [Press release]. Retrieved from https://www.thomsonreuters.com/en/press-releases/2016/july/thomson-reuters-announces-definitive-agreement-to-sell-its-intellectual-property-science-business.html

Times Higher Education. (2018). Europe teaching rankings. *Times Higher Education*. Retrieved from https://www.timeshighereducation.com/world-university-rankings/europe-teaching-rankings-2018-methodology

Waltman, L., van Eck, N., Visser, M., & Wouters, P. (2016). The elephant in the room: The problem of quantifying productivity in evaluative scientometrics. *Journal of Informetrics*, *10*(2), 671–674.

Wilsdon, J. (2015). *The metric tide: Independent review of the role of metrics in research assessment and management*. London: Sage Publications.

Wilsdon, J., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., & Wouters, P. (2017). *Next-generation metrics: Reponsible metrics and evaluation for open science* (Report of the European Commission Expert Group on Altmetrics). Brussels: European Commission.

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121–136.

Wouters, P. (1999). *The citation culture* (Doctoral dissertation). University of Amsterdam, Amsterdam, The Netherlands.

About the Author

Anne K. Krüger is Scientific Coordinator of the Interdisciplinary Research Group 'Changing Universities and Their Social Environment: Consequences for Academic Freedom?' at the Berlin Brandenburg Academy of Sciences and Humanities. Prior to this, she was Researcher at the Department of Science Studies at Humboldt University Berlin and leader of the research group 'Evaluation Practices in Science and Higher Education' at the German Centre for Higher Education Research and Science Studies (DZHW). In her current research, she focuses on evaluative information infrastructures and the sociology of (e)valuation.

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2020, Volume 8, Issue 2, Pages 68–71
DOI: 10.17645/pag.v8i2.2991

Commentary

Governance by Numbers: A Panopticon Reversed?

Julian Hamann

Leibniz Center for Science and Society, Leibniz University Hannover, 30159 Hannover, Germany;
E-Mail: julian.hamann@lcss.uni-hannover.de

Submitted: 8 March 2019 | Accepted: 16 March 2020 | Published: 9 April 2020

Abstract
This commentary is concerned with a specific form of power and discipline that is exerted through governance by numbers. Because of its many parallels to classical Foucauldian panopticism, governance based on numbers can be coined 'numerocratic panopticism.' Yet, going beyond similarities between classical and numerocratic panopticism, the commentary suggests three features specific to numerocratic panopticism that actually reverse characteristic traits of classical panopticism: In contrast to classical panopticism, numerocratic panopticism is multi-centered, non-spatial and open-purpose. Research on governance by numbers can benefit from a heuristic of panopticism if it considers both similarities and differences between classical and numerocratic panopticism.

Keywords
discipline; dispositif; Foucault; numerocracy; panopticon; power; quantification; surveillance

Issue
This commentary is part of the issue "Quantifying Higher Education: Governing Universities and Academics by Numbers" edited by Maarten Hillebrandt (University of Helsinki, Finland) and Michael Huber (University of Bielefeld, Germany).

1. Introduction

In the wake of New Public Management, public service sectors are increasingly governed by numbers. The cold, rational gaze of indicators, metrics and algorithms is supposed to increase productivity and efficiency and facilitate transparency and accountability. This new governance landscape also seizes the higher education sector: Governance based on numbers is developed and enforced by political stakeholders and media corporations and employed at different levels from the state to the university to individual departments. The contributions to this thematic issue survey this landscape. Krüger (2020), Hillebrandt (2020), and Kandiko Howson and Buckley (2020) illustrate how increasingly extensive and elaborate data infrastructures are becoming an end in themselves. The contributions by Dix, Kaltenbrunner, Tijdink, Valkenburg, and de Rijcke (2020) and Huber (2020) show how performance-based budgeting and quality assurance schemes direct universities, departments and researchers toward new objectives. Ringel, Brankovic, and Werron (2020) demonstrate how organizations ensure the ongoing production and promotion of rankings as a numerical observation of higher education institutions.

In the following, I will zoom in on the form of power and discipline that is exerted by governance based on numbers. Against the backdrop of Michel Foucault's socio-historical study on panopticism (Foucault, 1991), I will discuss different aspects of governance by numbers as panopticism.

2. Governance by Numbers as Panopticism

The contributions to this thematic issue create an overview of governance by numbers as an arrangement of discourses, devices, practices and infrastructures that facilitate the performance-oriented steering of higher education. This provides the opportunity to reflect again upon governance by numbers as a Foucauldian power-knowledge complex based on panoptical dispositifs (Foucault, 1977, p. 194). In doing so, my aim is to show how research on governance by numbers can benefit from a Foucauldian perspective on power and discipline, a perspective that is all but new to the study

of quantification (cf. Mennicken & Espeland, 2019) but which nonetheless brings issues to the fore that can complement the current contributions.

A panopticon is both a specific type of institutional building and a general system of control. Originally developed by the philosopher and social reformer Jeremy Bentham in the 18th century, panopticism received popular attention when Foucault used it as an analogy for the emergence of the modern disciplinary society (Foucault, 1991). Epitomizing a general principle that allows for the efficient exertion of power and discipline, the panopticon was the ideal metaphor for Foucault to explain how human complexities are ordered and individuals are inserted seamlessly into the social machinery. The classical panoptical architecture is realized in a circular building. The periphery of the rotunda is divided into cells with two windows: One that opens on the outside and allows the light to cross the cell from one side to the other, and another window that opens toward the center. In the center is a tower with windows that open toward the peripheral cells around it. The tower is occupied by an individual in task of surveying the separated, perfectly individualized cells. While each cell is constantly visible from the tower and the tower is visible from the cells, the inmates never know whether they are actually being observed at a specific point in time. The appeal that the panopticon had for contemporaries is (at least) twofold: First, there is no escape from the panoptic gaze. It can penetrate each cell. Second, power relations are automatized. They are upheld mechanically, without force, chains and dungeons.

Discussing governance by numbers as a form of panopticism, I join other scholars who have identified different forms of number-based panopticism. The increase of quantification and the centralization of the production, collection and analysis of numerical data have led to what has been described as statistical panopticism (Diaz-Bone, 2019). Based essentially on numerical information, statistical panopticism generates and accumulates huge masses of data, for example in the form of official statistics on unemployment or as data on health gathered by apps developed in insurance companies. The examples already convey that the potential of number-based panopticism has exponentially increased due to technological developments. Consequently, information technology in the business sector has been described as an information panopticon that records and displays human behavior and thus provides the computer age with a new degree of transparency (Zuboff, 1988). Although similar to statistical panopticism, an information panopticon is not necessarily characterized by huge masses of data. It is usually limited to the workplace and thus geared more directly towards the managerial control of workers. The academic equivalent of the information panopticon in the business sector are numerocratic techniques of scientometrics, which exercise a power that aims to govern large populations of academics through numbers and standards (Angermuller & van Leeuwen, 2018). Current digital information technologies have long potentiated the means of information panopticon. Digital archives from Google Scholar to academia.edu are deeply embedded in the social organization of the sciences. They form what has been coined a digital panopticon in which everybody can observe everybody else—and not least themselves (Angermüller, 2010).

Reflections on statistical, informational and digital panopticism have revealed important insights on governance based on numbers. Although it employs different foci, the literature illustrates that governance by numbers can be conceived as a panoptical dispositif of power and discipline. The literature shares the basic assumption that there are marked similarities between classical and contemporary panoptical dispositifs.

3. Some Specifics of Numerocratic Panopticism

The similarities between classical panopticism and governance based on numbers notwithstanding, I would like to stress three differences between the two panoptical dispositifs. This may specify the heuristic of panopticism and contribute to a clearer picture of what could be coined 'numerocratic panopticism,' a form of government in which the authority of numbers is exerted through panoptical power and discipline (cf. Angermuller & van Leeuwen, 2018).

The first difference between classical and numerocratic panopticism concerns the way in which the observer and the observed are arranged in relation to each other. In the classical panopticon, observed subjects are locked in separated cells around a central tower from which they are observed. Numerocratic panopticism reverses this positional arrangement of the observer and the observed in two ways. First, the new positional arrangement does not lock up the observed in a peripheral cell, but places them in the center and surrounds them with observers. There is no single disciplining gaze emanating from the center. Rather, academic subjects are now observed by a number of different indicators and metrics. While the revolutionary potential of the classical panopticon is based on its ability to "reduce the number of those who exercise power" (Foucault, 1991, p. 206), numerocratic panopticism reverses this logic and increases the number of observers. In this regard, a heuristic of panopticism could complement Huber's (2020) contribution on financial quantification in universities, which attends to various types of numbers that are produced for a multiplicity of audiences. The second reversal of the positional arrangement of the classical panopticon is that the aim of numerocratic panopticism is not to isolate the observed in separated cells. Instead, the observed subjects can see each other, indeed, they are supposed to mutually observe, compare and discipline each other. In the numerocratic panopticon, the disciplining gaze does not only emanate from multiple observers but also from other observed subjects, and, finally, from each subject observing itself. University rank-

ings, as analyzed by Ringel et al. (2020), are a prime example of this form of discipline.

Although Foucauldian panopticism is a general principle for the exertion of control and power, it has to be realized in spatial arrangements. Classical panopticism is dependent on a specific architectural form that has been adopted by disciplinary institutions like psychiatric asylums or penitentiaries. Governance based on numbers does not rely on such spatial constraints. Its realization neither depends on the spatial arrangements of buildings nor on the mutual physical presence of the observer and the observed in the same locality. Although indicators and metrics as well as the paper sheets and digital databases hosting them can have a material form, they are void of a specific spatial localization and do not require the individual to be "inserted in a fixed place" (Foucault, 1991, p. 197). To be sure, the formula for the spread of panopticism beyond enclosed spaces already lies in the panoptic principle itself and the spread of disciplinary institutions during the classical age. Yet, the reversal of the distinctively spatial arrangement of classical panopticism is only fully realized by omnipresent, non-spatial numerocratic panopticism that resembles a fluid network rather than an architectural form. The non-spatiality and omnipresence of governance by numbers is illustrated in the contributions by Krüger (2020) and Ringel et al. (2020) on data infrastructures and rankings, which are precisely located at the intersection of politics, economy, media and academia, interrelating different practices, organizations and research fields. A heuristic of panopticism could complement these contributions, revealing how panoptical dispositifs allow specific forms of power and discipline to permeate practices, organizations and fields.

As a third difference to classical panopticism, numerocratic panopticism is much more open regarding the specific purpose of observation and control. The ultimate rationale of the classical panopticon is "measuring, supervising and correcting the abnormal" in order "to train or correct individuals" (Foucault, 1991, pp. 199, 203). Crucially, this rationale is achieved because the observed can never be sure whether they are actually being observed, even though the tower is constantly visible. Numerocratic panopticism reverses this setting: First, subjects do know that they are being observed at any moment, that each action is tracked and filed. Yet, unlike the tower, the observing entity is not in plain sight but has diffused into algorithms, metrics and indicators. Second, it is much more difficult to pinpoint the purpose of individual observations. Observed subjects do not automatically know what each observer expects from them. What is more, some observations might bear no relevance at all for the observed. At the very least, it is an empirical question whether data collected by, for example, academia.edu has any impact on academics' everyday life. Contributions to this thematic issue emphasize this: They show that observers produce numbers without actually using them (Hillebrandt, 2020), that numbers are produced without an explicit purpose (Kandiko Howson et al., 2020; Krüger, 2020), and that the purpose of observations can be challenged and contested (Dix et al., 2020). A heuristic of panopticism could complement these contributions by raising questions about the disciplining effects of observations that are more open regarding their specific purpose. A Foucauldian perspective suggests that this openness is precisely the foundation for the self-discipline of subjects that is already implied in the panopticon and fully developed in governmentality (Foucault, 2010).

4. Conclusion

Although classical and numerocratic panopticism share many similarities, I have proposed three features that distinguish them. Future research on governance by numbers can benefit from a heuristic of panopticism if it considers not only similarities, but also the differences between classical and numerocratic panopticism. First, numerocratic panopticism is multi-centered. Instead of a single observing entity, numerocratic panopticism is characterized by a multitude of different observers that have moved to the periphery. They surround the observed that is now placed in the center. Second, numerocratic panopticism is non-spatial. It is an omnipresent, fluid network that does not rely on specific spatial arrangements but permeates practices, organizations and fields. Third, numerocratic panopticism is open-purpose. The observed is not aware of the exact purposes of any individual observation, and some observations may be entirely insignificant to the observed. Crucially, these three distinct features of numerocratic panopticism reverse the respective features of the classical panopticon. This suggests that governance by numbers is not only an epitome of classical panopticism but, at least in some key aspects, also a panopticon reversed.

Acknowledgments

This comment has benefitted from discussions with Stefan Beljean, Rainer Diaz-Bone, Frank Meier and Thorsten Peetz during a meeting of the DFG research network "Towards a Society of Valuation?" I am also thankful for constructive feedback from the two academic editors and helpful comments from my colleagues at LCSS.

Conflict of Interests

The author declares no conflict of interests.

References

Angermüller, J. (2010). Wissenschaft zählen: Regieren im digitalen Panoptikum [Counting science: Governance in the digital panopticon]. In L. Hempel, S. Krasmann, & U. Bröckling (Eds.), *Sichtbarkeitsregime: Überwachung, Sicherheit und Privatheit*

im 21. Jahrhundert [Regimes of visibility: Surveillance, security and privacy in the 21st century] (pp. 174–190). Wiesbaden: Springer VS.

Angermuller, J., & van Leeuwen, T. N. (2018). On the social uses of scientometrics: The quantification of academic evaluation and the rise of numerocracy in higher education. In R. Scholz (Ed.), *Quantifying approaches to discourse for social scientists* (pp. 89–119). London: Palgrave Macmillan.

Diaz-Bone, R. (2019). Statistical panopticism and its critique. *Historical Social Research*, *44*(2), 77–102.

Dix, G., Kaltenbrunner, W., Tijdink, J., Valkenburg, G., & de Rijcke, S. (2020). Algorithmic allocation: Untangling rival considerations of fairness in research management. *Politics and Governance*, *8*(2), 15–25.

Foucault, M. (1977). The confession of the flesh. In C. Gordon (Ed.), *Power/knowledge: Selected interviews & other writings, 1972–1977, by Michel Foucault* (pp. 194–228). New York, NY: Pantheon Books.

Foucault, M. (1991). *Discipline and punish: The birth of the prison*. New York, NY: Vintage Books.

Foucault, M. (2010). *The government of the self and others: Lectures at the Collège de France 1982/83*. New York, NY: Palgrave.

Hillebrandt, M. (2020). Keeping one's shiny Mercedes in the garage: Why higher education quantification never really took off in Germany. *Politics and Governance*, *8*(2), 48–57.

Huber, M. (2020). Steered by numbers: How quantification differentiates the reform of a German university. *Politics and Governance*, *8*(2), 26–35.

Kandiko Howson, C., & Buckley, A. (2020). Quantifying learning: Measuring student outcomes in higher education in England. *Politics and Governance*, *8*(2), 6–14.

Krüger, A. K. (2020). Quantification 2.0? Bibliometric infrastructures in academic evaluation. *Politics and Governance*, *8*(2), 58–67.

Mennicken, A., & Espeland, W. (2019). What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, *45*(2019), 223–245.

Ringel, L., Brankovic, J., & Werron, T. (2020). The organizational engine of rankings: Connecting "new" and "old" institutionalism. *Politics and Governance*, *8*(2), 36–47.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. New York, NY: Basic Books.

About the Author

Julian Hamann is Postdoc at the Leibniz Center for Science and Society, Leibniz University Hannover, Germany. He is a Sociologist working in sociology of science and higher education research and at the crossroads of the sociologies of knowledge, culture and inequality. Recent research has been published in *Social Studies of Science* and *American Journal of Cultural Sociology*.

Politics and Governance
Open Access Journal | ISSN: 2183-2463
Volume 8, Issue 2 (2020)

Quantifying Higher Education: Governing Universities and Academics by Numbers

Editors

Maarten Hillebrandt and Michael Huber

COGITATIO

Politics and Governance, 2020, Volume 8, Issue 2
Quantifying Higher Education: Governing Universities and Academics by Numbers

Published by Cogitatio Press
Rua Fialho de Almeida 14, 2º Esq.,
1070-129 Lisbon
Portugal

Academic Editors
Maarten Hillebrandt (University of Helsinki, Finland)
Michael Huber (University of Bielefeld, Germany)

Available online at: www.cogitatiopress.com/politicsandgovernance

Table of Contents

Politics and Governance (ISSN: 2183-2463)

Politics and Governance is an innovative new offering to the world of online publishing in the Political Sciences. An internationally peer-reviewed open access journal, Politics and Governance publishes significant, cutting-edge and multidisciplinary research drawn from all areas of Political Science.

www.cogitatiopress.com/politicsandgovernance

COGITATIO